


2017

L2 Learners' Perceptions and Preferences of Automated Corrective Feedback

Nawaf Alsallami
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>

 Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), [English Language and Literature Commons](#), [Linguistics Commons](#), and the [Other Languages, Societies, and Cultures Commons](#)

Recommended Citation

Alsallami, Nawaf, "L2 Learners' Perceptions and Preferences of Automated Corrective Feedback" (2017). *Graduate Theses and Dissertations*. 15244.
<https://lib.dr.iastate.edu/etd/15244>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

L2 learners' perceptions and preferences of automated corrective feedback

by

Nawaf Alsallami

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Applied Linguistics and Technology

Program of Study Committee:

Volker H. Hegelheimer, Major Professor

Carol A. Chapelle

Evgeny Chukharev-Hudilainen

Geoffrey F. K. Sauer

Gary D. Phye

Xiaoqiu Huang

The student author and the program of study committee are solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

Copyright © Nawaf Alsallami, 2017. All rights reserved.

To my mother Noor and father Naeem
for raising me to be what I am now,
for their patience, care, and never-ending love,
and for the many sacrifices they made to see me successful.

To my wife Sabah
for her enormous and indispensable support throughout my educational journey,
for her patience in the years she spent away from her family,
for heroically taking care of our children so that I focus on my studies,
and for her sincere and passionate love that kept me going through hard times.

To my children Albaraa, Sura, and Sarah
for being the joy of my life and the coolness of my eyes
and for always making me a proud father.

I will forever be indebted to you!

TABLE OF CONTENTS

LIST OF FIGURES	x
LIST OF TABLES	xi
TERMINOLOGY	xii
ACKNOWLEDGEMENT	xiv
ABSTRACT	xv
CHAPTER 1. INTRODUCTION	1
1.1. Statement of the Problem	3
1.2. Purpose of the Study	4
1.3. Research Questions	4
1.4. Significance of the Study	5
CHAPTER 2. LITERATURE REVIEW	6
2.1. Overview of Corrective Feedback	6
2.1.1. Theoretical underpinnings.	6
2.1.1.1. Conscious vs. unconscious language learning.	7
2.1.1.2. Focus on form.	8
2.1.2. Effectiveness of CF in promoting L2 development.	8
2.1.3. Mediating factors of corrective feedback.	9
2.1.3.1. Feedback explicitness.	9
2.1.3.2. Feedback medium.	10
2.1.3.3. Feedback focus.	11

2.2. Frequency and Learner Uptake of Corrective Feedback	12
2.2.1. Frequency and uptake of oral CF.	13
2.2.1.1. Clarification requests.	13
2.2.1.2. Elicitation.	14
2.2.1.3. Explicit correction.	15
2.2.1.4. Metalinguistic feedback.	17
2.2.1.5. Recasts.	17
2.2.1.6. Repetition.	18
2.2.1.7. Translation.	18
2.2.2. Frequency and uptake of written CF.	18
2.3. Corrective Feedback and L2 Development: Meta-analysis	20
2.3.1. Method.	21
2.3.1.1. Identification of primary studies.	21
2.3.1.2. Inclusion and exclusion criteria.	22
2.3.1.3. Coding.	23
2.3.1.3.1. Corrective feedback strategies.	23
2.3.1.3.2. Timing of posttests.	25
2.3.2. Analysis.	25
2.3.2.1. Meta-analysis models.	26
2.3.2.2. Calculation of effect sizes.	26
2.3.2.3. Sample size inflation.	27

2.3.2.4. Outliers.	27
2.3.2.5. Publication bias.	28
2.3.3. Results.	28
2.3.3.1. Oral CF and L2 development.	28
2.3.3.1.1. Effect of oral corrective feedback.	31
2.3.3.2. Written CF and L2 development.	33
2.3.3.2.1. Effect of written corrective feedback.	33
2.3.4. Discussion.	34
2.3.4.1. Oral CF discussion.	34
2.3.4.2. Written CF discussion.	37
2.3.5. Conclusion.	39
2.4. CF in CALL	39
2.5. Use of Sociocultural Theory in CALL and CF	42
2.5.1. Overview of Sociocultural Theory.	42
2.5.1.1. Mediation.	43
2.5.1.2. Social learning.	44
2.5.1.3. Genetic analysis.	44
2.5.2. Use of Sociocultural Theory in CALL.	44
2.5.2.1. Mediation and CALL.	45
2.5.2.2. Social learning in CALL.	45
2.5.2.3. Genetic analysis in CALL.	47

2.5.3. Use of Sociocultural Theory in CF.	47
2.5.3.1. Graduated CF.	48
2.6. Perceptions and preferences of CF and its strategies	50
2.7. Research Questions	55
CHAPTER 3. METHODOLOGY	57
3.1. Design	57
3.2. Target Structure	58
3.3. Participants	59
3.4. Materials	59
3.4.1. AWE Tool	60
3.4.1.1. Text editor.	61
3.4.1.2. Analysis module.	61
3.4.1.3. Feedback module.	62
3.4.1.4. Logging module and survey module.	64
3.4.2. Video-stimulated recall.	64
3.4.3. Survey.	65
3.4.4. Semi-structured interview.	66
3.5. Data Collection Procedure	66
3.6. Data Analysis	68
3.6.1. Research Question 1.	68
3.6.2. Research Question 2.	69
3.6.3. Research Question 3.	70

3.7. Chapter Summary	70
CHAPTER 4. RESULTS AND DISCUSSION	72
4.1. Factor Analysis Results	73
4.2. Research Question 1: CF Perception and Timing and Frequency Preferences	75
4.3. Research Question 2: Perceptions and Preferences of CF Strategies	91
4.3.1. Identification	95
4.3.2. Direct correction	99
4.3.3. Metalinguistic CF	103
4.3.4. Graduated CF	107
4.4. Research Question 3: Past Experience and CF Preferences and Perceptions	119
4. 5. Chapter Summary	121
CHAPTER 5. CONCLUSION	123
5.1. Summary of Major Findings	124
5.2. Implications and Recommendations	125
5.3. Limitations of the Study	127
5.4. Conclusion	129
REFERENCES	130
APPENDIX A. VIDEO-STIMULATED RECALL QUESTIONS	150
APPENDIX B. SURVEY	151
APPENDIX C. INTERVIEW QUESTIONS	155

APPENDIX D. REPOSITORIES, PUBLICATIONS, AND KEYWORDS USED IN LITERATURE SURVEY	156
1. List of academic publication repositories searched for primary CF studies	156
2. List of peer-reviewed journals searched for primary CF studies	156
3. Search keys used in searching for primary CF studies	157
APPENDIX E. FORTY-FOUR CF STUDIES QUALIFIED FOR A META-ANALYSIS	158
1. Twenty-Five Oral CF Studies Qualified for a Meta-Analysis	158
2. Nineteen Written CF Studies Qualified for a Meta-Analysis	159
APPENDIX F. META-ANALYSIS RESULTS	161
APPENDIX G. PERCEPTIONS AND PREFERENCES OF CF	167
APPENDIX H. PERCEPTIONS OF GRADUATED CF AND WTUTOR	173
APPENDIX I. WTUTOR SCREENSHOTS	178
1. Initial state of <i>wTutor</i> editor	178
2. <i>wTutor</i> reporting error	179
3. <i>wTutor</i> explaining error type	180
4. <i>wTutor</i> evaluating knowledge about error type	181
4.1. wTutor prompting user to select error type	181
4.2. wTutor prompting user to correct sample sentence	182
5. <i>wTutor</i> ending dialog	183
5.1. wTutor prompting user to correct similar errors	183
5.2. State of editor after wTutor ends dialog	184

APPENDIX J. PLOT FOR EXPLORATORY FACTOR ANALYSIS	185
APPENDIX K. R SCRIPT OF DESCRIPTIVE STATISTICS AND EXPLORATORY FACTOR ANALYSIS	186

LIST OF FIGURES

Figure 1. Normal probability Q-Q plot for oral CF studies	29
Figure 2. Funnel plot of precision by standard difference in means of oral CF studies	31
Figure 3. Normal probability Q-Q plot for written CF studies	34
Figure 4. Funnel plot of precision by standard difference in means of written CF studies	35
Figure 5. Effects of oral CF strategies over time	36
Figure 6. Effects of written CF strategies over time	38
Figure 7. A screenshot of <i>wTutor</i>	61
Figure 8. Screeplot showing the number of optimal factors to extract from surveys	73

LIST OF TABLES

Table 1. Occurrence Frequency Rate of Different Strategies of Oral Corrective Feedback	15
Table 2. Uptake Frequency Rate of Different Strategies of Oral Corrective Feedback	16
Table 3. Meta-analysis Coding Scheme	24
Table 4. Outlier Oral CF Studies	30
Table 5. Effects of Oral Corrective Feedback	32
Table 6. Effects of Written Corrective Feedback	35
Table 7. Usefulness Perception of Six CF Strategies	53
Table 8. Description of Dialog States	62
Table 9. Results of Exploratory Factor Analysis	74
Table 10. Summary of Responses to CF Perception and Preference Statements on a 5-Point Likert Scale (n=30)	76
Table 11. Summary of Responses to Timing and Frequency Preference Statements on a 5-Point Likert Scale (n=30)	76
Table 12. Summary of Responses to CF Strategies' Perception and Preference Statements on a 5-Point Likert Scale (n=30)	93
Table 13. Summary of Responses to CF Experience Statement on a 5-Point Likert Scale (n=30)	120

TERMINOLOGY

AWE Automated Writing Evaluation

AWE is the process of analyzing written essays, identifying formal, functional and structural errors and providing appropriate feedback by the computer.

CALL Computer-Assisted Language Learning

CALL is second language learning that involves using computer technologies such as personal computers and smartphones.

CF Corrective Feedback

CF is feedback about language errors in the speech or writing of second language learners.

CMC Computer-Mediated Communication

CMC is communication between individuals through computer technologies such as Skype and online forums.

EFL English as Foreign Language

EFL refers to English when used in settings where it is not spoken as a native language such as in Iraq and Saudi Arabia.

ESL English as Second Language

ESL refers to English when used in settings where it is spoken as a native language such as in the United States and the United Kingdom.

FonF Focus on Form

FonF refers to “occasional shift of attention to linguistic code features - by the teacher and/or one or more students - triggered by perceived problems with

comprehension or production [during meaning-oriented tasks and interactions]”
(Long & Robinson, 1998, p. 23).

L2 Second Language

L2 refers to a language that is not native to individuals who are using and learning it.

SCT Sociocultural Theory

“Vygotsky’s sociocultural theory of human learning describes learning as a social process and the origination of human intelligence in society or culture. The major theme of Vygotsky’s theoretical framework is that social interaction plays a fundamental role in the development of cognition. Vygotsky believed everything is learned on two levels. First, through interaction with others, and then integrated into the individual’s mental structure.” (UNESCO, 2003)

ACKNOWLEDGEMENT

Undoubtedly, this project reached fruition with the help and support of numerous individuals but only after the indispensable support, guidance, and will of Allah, the Most Merciful and Compassionate. I sincerely thank Allah for making my dream come true. I also thank the following people for their assistance and support throughout my project:

- to my major professor, Volker Hegelheimer, for his guidance and encouragement;
- to the members of my POS committee for their constructive feedback;
- to the ESL instructors and administrators of the Intensive English and Orientation Program at Iowa State University and DMACC Ames Career Academy Hunziker Center for allowing me to work with their ESL students, especially Ellie Caldwell for her tremendous support;
- to all participants who joined my project;
- and to the University of Tabuk for granting me a full scholarship that enabled me to work on this project in the first place.

ABSTRACT

Form-focused automated Corrective Feedback (CF) is widely used in general-purpose and specialized writing software and research indicates a positive effect for automated CF on second language (L2) learning (AbuSeileek, 2013; AbuSeileek & Abualsha'r, 2014). As any educational practice, preferences and perceptions of learners are expected to influence how automated CF is used by L2 learners (Amrhein & Nassaji, 2010; Brown, 2009; Schulz, 2001). However, preferences and perceptions of automated CF are poorly understood due to paucity of relevant research. This study contributes to this line of investigation by exploring three pertinent topics. First, it explores L2 learners' preferences and perceptions of automated CF and four different CF strategies: identification, direct correction, metalinguistic CF, and graduated CF. Second, it examines learners' preferences between different CF timings and frequency choices. Third, it explores learners' past experience with AWE tools and investigates if past Automated Writing Evaluation (AWE) experience affects learners' preferences and perceptions of automated CF.

To accomplish these objectives, the present study surveyed and interviewed 30 learners at an intermediate to advanced English as Second Language (ESL) proficiency level. It calculated descriptive statistics of the surveys and employed exploratory factor analysis to identify the underlying relationships between different variables measured by the survey. For interview analysis, it employed a grounded theory approach to identify major concerns and perceptions of automated CF not accounted for in the survey. Results revealed a strong preference for direct correction followed by metalinguistic CF, identification, and graduated CF respectively. Factor analysis identified a close association between clarity and usefulness perceptions and preferences for CF strategies, between comprehensive CF and direct correction and between the frequency of

AWE use and identification. The interviews revealed two major concerns with potential influence on preferences for CF strategies: time and learning. Based on preference data, the time required to use CF successfully for error correction is a more important factor for most learners than learning from CF. In other words, CF strategy preferences appear to be mainly shaped by the time factor. This dissertation concludes with specific implications of these findings for developers of AWE tools and L2 educators. Specifically, developers should be mindful of the wide range of concerns that shape L2 learners' preferences and perceptions of CF in order to design and deliver CF that is timely, desirable, and positively perceived by L2 learners. Furthermore, L2 educators should exert the effort to mitigate L2 learners' concerns that undermine the value of CF qualities and strategies that were empirically proven to be effective for L2 development.

CHAPTER 1. INTRODUCTION

Language is essentially a tool for expressing meaning. This tool is strictly codified into a set of rule-governed systems: semantic, syntactic, phonological, and pragmatic. Successful use of language must conform to the conventions of these systems (Finegan, 2012). Divergence from these conventions often leads to communication problems and poor comprehensibility of intended messages. For this reason, error correction has played a pivotal role in Second Language (L2) education.

Error correction has garnered much attention due to the typical divergence of developing learners' interlanguage from target language (Ellis, 2009b). This error correction practice has been approached from different perspectives depending on how L2 development and acquisition is viewed. In a naturalistic view of L2 development, language accuracy is promoted through extensive exposure to meaningful and contextualized language input without direct error correction (Krashen, 1981). However, research findings suggest that language exposure is not sufficient for reaching advanced competency in all language areas especially in the area of language accuracy (Swain & Herron, 1987). The interactionist view of L2 development argues that Focus-on-Form (FonF) is required for reaching higher levels of language accuracy (Long, 1988, 1991). FonF involves correcting formal language errors incidentally as they occur during meaning-oriented interactions in the hope of raising learners' awareness of those errors and pushing them to internalize correct formal rules.

Scores of empirical studies have investigated the effectiveness of FonF in promoting L2 development and attempted to identify how its effectiveness is mediated by various components such as comprehensible input, comprehensible output, Corrective Feedback (CF), noticing, and uptake. These components are closely related and oftentimes occur together making it

challenging to pinpoint the exact contribution of each component. However, the starting point for investigating the effectiveness of FonF is CF. CF on a formal error may lead to learners' noticing of the error, which in turn may be followed by further interactions where comprehensible input and output are delivered and uptake is realized. CF is a response to errors in learner utterances consisting of "(a) an indication that an error has been committed, (b) provision of the correct target language form, or (c) metalinguistic information about the nature of the error, or any combination of these" (Ellis, Loewen, & Erlam, 2006, p. 340).

Research has identified a positive effect for both traditional and automated written CF on L2 writing development and varying effect for different CF strategies (AbuSeileek & Abualsha'r, 2014; Kang & Han, 2015; Russell & Spada, 2006). However, despite favorable research findings and common use of CF in L2 settings, learners' preferences and perceptions of CF and its different strategies have not been fully understood and investigated. In recent years, automated CF became widely used in L2 software, including Automated Writing Evaluation (AWE) tools such as ETS Criterion®, MY Access!®, and generic word processors. However, very little is known about how L2 learners perceive the different qualities and strategies of automated CF. Preferences and perceptions of CF are important factors determining how learners interact and make use of CF, which in turn impact its effectiveness (Amrhein & Nassaji, 2010; Brown, 2009; Schulz, 2001).

CF takes on a number of qualities related to timing, directness, comprehensiveness, technicality, and graduation. Variation in the directness and technicality of CF produces three main strategies: identification, direct correction, and metalinguistic explanation. Identification involves visually highlighting or underlining errors without supplying a corrected version of the erroneous text or providing a technical explanation of the error; direct correction involves

supplying learners with a corrected version of their erroneous text; and metalinguistic explanation involves technically explaining the type and cause of the error. CF could also involve a combination of these strategies such as a metalinguistic explanation with direct correction. On the other hand, variation in the graduation quality produces graduated CF that involves prioritizing other CF qualities (i.e., directness and technicality) based on learners' Zone of Proximal Development (ZPD) and delivered through a dialog (Aljaafreh & Lantolf, 1994). ZPD is "the difference between what an individual is able to accomplish functioning alone (i.e., actual level of development) and what that person is capable of in collaboration with other, more expert, individuals (i.e., potential level of development)" (Lantolf & Aljaafreh, 1995, p. 619). Dialogic interaction helps in determining a learner's ZPD and delivering appropriate CF that promotes self-reliance. Graduated CF proved very effective in promoting L2 development in non-CALL settings (Erlam et al., 2013; Rassaei, 2014). However, research introducing automated graduated CF and exploring learners' preferences and perceptions of automated graduated CF is non-existent.

1.1. Statement of the Problem

The paucity of automated CF research is evident despite the significant and innovative potential of technology in facilitating the provision and personalization of CF. Research is particularly lacking on learners' preferences and perceptions of automated written CF. In recent years, a number of AWE tools have been developed to assist L2 learners to produce more accurate, well-structured, and genre-compliant text. CF is the main mechanism used by these tools to achieve this objective. Learners differ in their preferences and perceptions of these different CF qualities and strategies. Their preferences and perceptions should be considered when designing, developing, and using AWE tools to minimize friction between what learners

expect and want and what these tools provide. Furthermore, learners' input should be sought for any new CF strategy in order to maximize its appropriateness, relevance, and clarity. Graduated CF is one such strategy that has not been used in AWE tools so far.

1.2. Purpose of the Study

Two problems should be evident from the discussion above. First, L2 learners' preferences and perceptions of automated written CF and different automated CF strategies are poorly understood. Second, automated graduated CF has not been offered in AWE tools so far, and therefore learners' preferences and perceptions of it remain unexplored. In light of this, this study explores L2 learners' preferences and perceptions of automated CF and different automated CF strategies: identification, direct correction, metalinguistic CF, and graduated CF. Specifically, it examines learners' preferences of different automated CF timings and frequency choices. It explores how useful learners perceive automated CF for learning and how desirable, clear, and useful learners perceive different automated CF strategies. Lastly, it examines if past AWE experience and usage frequency have any bearing on learners' preferences and perceptions. These objectives are approached using an expansion-based mixed method design where multiple methods are employed to expand our understanding of L2 learners' preferences and perceptions.

1.3. Research Questions

The present study seeks to answer the following questions:

1. How do intermediate to advanced L2 learners perceive automated written CF and what are their preferred CF frequency and timing?
2. How do intermediate to advanced L2 learners perceive different automated written CF strategies in regards to their desirability, clarity, and usefulness?

3. What effects do usage frequency and past experience with AWE tools have on intermediate to advanced L2 learners' preferences and perceptions of automated CF and its different strategies?'

1.4. Significance of the Study

Improving the state of automated CF delivery requires integrating L2 learners' feedback in the design and development process of AWE tools. L2 learners do not necessarily hold the same view as software designers on what constitutes desirable, clear, and useful CF. In order to retain L2 learners' interest and boost their confidence in using automated CF, their preferences and perceptions should play an integral part of the design and development cycle. However, it is expected that not all learners' preferences and perceptions of automated CF agree with research findings about effective CF qualities and strategies. In this case, identifying these preferences and perceptions would help in anticipating, addressing, and potentially treating any misconceptions and negative attitudes. The present study contributes to this area of investigation by examining L2 learners' preferences and perceptions of different automated CF qualities and strategies. It sampled participants from the intermediate to advanced adult L2 learner population in the United States; introduced them to automated graduated CF by means of a custom-made proof-of-concept AWE tool; and surveyed and interviewed them about their perceptions and preferences of different qualities and strategies of automated CF.

CHAPTER 2. LITERATURE REVIEW

This literature review is organized into six main parts. The first part presents an overview of Corrective Feedback (CF) with a discussion of the theoretical underpinnings of CF, the effectiveness of CF in promoting L2 development, and factors mediating the effectiveness of CF. The second part provides a review of descriptive studies that examined the frequency of CF in L2 classrooms and expert-learner communications. It also reviews studies that examined the frequency by which uptake occurs in relation to different CF strategies. Uptake refers to learners' responses to CF, which can range from simple acknowledgement to modification of erroneous output. The third part systematically reviews experimental and quasi-experimental studies that investigated the effects of CF and different CF strategies on morphosyntactic L2 development. The fourth part reviews studies that examined the effectiveness of CF in Computer-Assisted Language Learning (CALL) settings. The fifth part reviews studies that explored the effectiveness of sociocultural-based graduated CF. The sixth and last part reviews studies that examined how CF is perceived and what CF qualities are preferred by L2 writers.

2.1. Overview of Corrective Feedback

2.1.1. Theoretical underpinnings.

CF is grounded in two theoretical foundations related to the role of conscious and focus on form in L2 development. Conscious language learning involves awareness of linguistic forms and structures brought by self-initiated reflection or other-initiated prompts. In contrast, Focus on Form (FonF) is a subset of conscious language learning and involves learners responding to other-initiated signals about erroneous forms and structures in their own spoken or written language during meaning-oriented activities (Long, 1988) These two foundations of CF are discussed next.

2.1.1.1. Conscious vs. unconscious language learning.

The nineteenth and twentieth centuries have witnessed the rise of many psychological schools of thought, the most prominent of which was behaviorism. Behaviorist theorists place all focus on observable behaviors and thus discredit mental processes, including consciousness. They claim that “consciousness cannot be scientifically investigated... [and b]ecause conscious experience is subjective, external observation is impossible, and introspection is discounted as being 'notoriously' unreliable” (Schmidt, 1990, p. 130). This view of the role of conscious in observable behavior had taken its toll on SLA theories with some claiming that “language learning is essentially unconscious” (Schmidt, 1990, p. 129) and others claiming that language learning is exclusively a process of habit formation (Mowrer, 1960). One of the prominent voices of this view of learning is Krashen, who drew a distinction between subconscious acquisition and conscious learning, which he regards as not useful in un-monitored language production and comprehension (1981). From this view of language learning sprang language teaching methods that gave little to no attention to formal language learning and error correction such as the Audiolingual, Natural, and Communicative approaches. These language teaching methods seek to replicate the conditions of L1 acquisition in L2 learning contexts in the hope of activating L1 acquisition processes in L2 learning (Krashen & Terrell, 1983; Krashen, 1985).

However, this view of limited to no role of consciousness in L2 learning has not gone uncontested for long especially with the emergence of empirical data that found a prominent role for formal language teaching in L2 learning. First, a shift has occurred in the cognitive psychology view of subjective awareness that came to the conclusion that “subjective awareness is [not] epiphenomenal, and... that learning without awareness is impossible” (Schmidt, 1990, p. 131). This shift has reverberated within the SLA research community with many schools of

thought emerging as a consequence, which give at least some value to formal teaching and conscious learning. Second, empirical studies (Doughty & Williams, 1998; Spada & Lightbown, 1993) into "Communicative, Natural, or Immersion L2 and FL programs demonstrated significant shortcomings in the accuracy of their [learners'] language" (Ellis, 2011, p. 36). Some argue that explicit focus on form and conscious-raising activities are required with L2 learners in order to improve their L2 accuracy (Ellis, 2007a; Rutherford & Sharwood Smith, 1985).

2.1.1.2. Focus on form.

Focus on Form (FonF) is a term first introduced by Long (1988) to refer to “occasional shift of attention to linguistic code features - by the teacher and/or one or more students - triggered by perceived problems with comprehension or production” (Long & Robinson, 1998, p. 23) during meaning-oriented tasks and interactions. It stands in the middle of the form-meaning continuum where at one pole stands exclusively meaning and communicative oriented L2 instruction and at the other stands discrete, isolated L2 forms instruction. FonF is built on the premise that attention to form is required for the acquisition of some L2 morphosyntactic structures and that FonF is most effective when learners are aware of both meaning and form. Effectiveness of FonF depends on such factors as timing, duration, target forms, setting, and learner age (Robinson, 2012). FonF can be realized through input enhancement and CF. Studies of the role of conscious in L2 development underscore the importance of CF in improving L2 accuracy.

2.1.2. Effectiveness of CF in promoting L2 development.

Fueled by a renewed interest in conscious learning, CF has been extensively studied in the last three decades (See Norris and Ortega (2000) and Ellis et al. (2006) for a meta-analysis of CF studies). Specifically, interest in empirically studying CF has been significantly increased as

a result of Truscott's (1996) strong claim that CF is not only ineffective but even harmful for L2 learning. This claim has generated substantial theoretical and empirical research into CF.

Theoretically, some researchers have undertaken the task of counter-arguing why CF should have a place in L2 education (e.g., Ferris, 1999). While others have undertaken the task of empirically putting the question to the test and investigating what effect if any CF has on L2 learning (e.g., Faqeih, 2012; Frear & Chiu, 2015; Jiang & Xiao, 2014; Kartchava & Ammar, 2014b; Li, 2010a, 2013; Monteiro, 2014; Mubarak, 2013; Parreno, 2013; Profozic, 2013; Rahimi & Zhang, 2013; Shintani & Ellis, 2013; Stefanou & Révész, 2015; Yilmaz, 2013).

This diverse set of empirical CF studies has identified a moderate to large durable positive effect for CF on L2 development across different modes (oral and written), contexts (laboratory and classroom), and degrees of explicitness (Ellis et al., 2006; Li, 2010b; Lyster & Saito, 2010; Russell & Spada, 2006). These findings strongly refute claims of no benefits or harmful consequences of CF. Undoubtedly, CF has a secure place in L2 education. As a consequence, the focus has moved to more specific questions about how different contextual and learner factors mediate the effect of CF on L2 development. Among the factors investigated are explicitness, mode, and focus of feedback. These factors are examined next.

2.1.3. Mediating factors of corrective feedback.

2.1.3.1. Feedback explicitness.

Explicit feedback refers to corrective moves that overtly declare a given utterance is ill-formed while implicit feedback refers to corrective moves that avoid such overt declaration by means of supplying positive evidence (e.g., recasts) and/or interactional moves (e.g., clarification requests) (Carroll & Swain, 1993). Positive evidence provides learners with the corrected versions of their erroneous utterances and leaves the task of recognizing its corrective purpose to

learners. On the other hand, interactional moves such as clarification requests and repetitions draw learners to their utterances without informing them what is wrong with them.

In writing, CF is invariably explicit by nature due to the absence of some interactional moves found in oral interaction that can be used to implicitly provide and/or request correction (Sheen, 2010a). For example, direct correction of written errors involves identifying the errors and supplying corrected versions, which makes direct correction very unlikely to be misunderstood for meaning- or communication-oriented feedback. This is true for other written CF strategies such as identification and metalinguistic explanation. In contrast, CF in oral interactions can be explicit or implicit. The corrective purpose of recasts and repetition, for instance, can be misunderstood because they are also used for meaning- or communication-oriented feedback. Explicit oral CF strategies such as metalinguistic explanation and explicit correction, on the other hand, are very clear about their corrective purpose. A number of studies found a larger effect size for explicit feedback while implicit feedback had a more durable effect (Carroll et al., 1992; Carroll & Swain, 1993; Li, 2010a). One plausible explanation for the observed advantage of explicit feedback over implicit feedback is that the former is more noticeable by learners than implicit feedback. Implicit feedback can be easily misunderstood for a meaning-oriented or communicative interactional move especially by lower proficiency L2 learners.

2.1.3.2. Feedback medium.

The written and oral media of feedback delivery have their own unique qualities that may have differential impacts on the effectiveness of CF. The oral medium demands more economic and selective use of attentional resources than the written medium due to shorter available processing time in the former (Schmidt, 2001). This means that more attention is allocated to

communicative and meaningful language features than less communicative and less meaningful formal features in oral performance (VanPatten, 1990). Thus, it is expected that the oral medium offers less opportunity than the written medium for noticing formal issues and as a consequence may benefit less from CF. However, no significant differences have been empirically detected between the effectiveness of CF in the two media (Sheen, 2010a). Both oral and written CF have a large effect size on L2 development (Russell & Spada, 2006). Any non-significant advantage observed for written CF over oral CF could be attributed to the more explicit nature of written CF (Sheen, 2010a).

2.1.3.3. Feedback focus.

CF can be of two types: focused and unfocused. Focused CF targets specific formal errors while unfocused CF equally targets all formal errors. Both focused and unfocused CF can be either explicit or implicit. Most L2 teachers do not focus on a single error type because they want to improve the overall accuracy of learners' utterances or writings. For this reason, CF studies that do not focus on a single error type can determine how effective unfocused teacher CF is in promoting L2 accuracy. On the other hand, focused CF studies help in determining the effect of CF in promoting L2 development because they control for variation in the learnability of different linguistic structures. For instance, focused CF studies could determine if the development of English as Second Language (ESL) past simple benefits less or more from CF compared with ESL present perfect. Few studies (Ellis et al., 2008; Frear & Chiu, 2015; Sheen et al., 2009) have directly compared the effectiveness of focused and unfocused CF and what studies do exist show conflicting results.

Ellis et al. (2008) and Frear and Chiu (2015) found comparable effects for focused and unfocused CF, but their findings are overshadowed by methodological limitations of small

sample size, unequal group performance on the pretests, short treatment session(s), and use of different CF strategies (direct vs. indirect CF). In contrast, Sheen et al. (2009), with a larger sample size (n=80) and a longer treatment (2 weeks), found a larger effect for focused CF compared with unfocused CF. However, Sheen et al. (2009) conducted the study in an ESL context unlike the English as a Foreign Language (EFL) contexts of the former two studies. Therefore, it is unclear if differences in observations can be attributed to context differences. Unfortunately, no definite conclusion can be reached concerning the differential effect of focused and unfocused CF types based on the studies addressing this topic so far.

2.2. Frequency and Learner Uptake of Corrective Feedback

The medium of instruction, whether written or spoken, influences the types of strategies that can be employed for feedback delivery. Therefore, some strategies of feedback delivery are unique to a particular medium of instruction and other strategies are common between both media of instruction. Oral CF can be realized through seven strategies: clarification requests, elicitation, explicit corrections, metalinguistic feedback, recasts, repetitions, and translation (Lyster & Ranta, 1997; Panova & Lyster, 2002; Suzuki, 2004). On the other hand, written CF can be realized through five strategies: direct correction, identification (indicating an error exists without providing correction), concordance feedback, metalinguistic CF, and reformulation (Ellis, 2009a). These strategies are defined in the sections that review each individual strategy below.

This section reviews studies that examined the frequency of different CF strategies in oral and written L2 interactions and their relation with subsequent learner uptake. Uptake refers to learner's responses to CF, which may constitute a repair attempt or just an acknowledgment of

CF (Lyster & Ranta, 1997). Uptake is considered a precursor of noticing which in turn is considered a precursor of L2 development (Loewen, 2005; Mackey, 2006).

2.2.1. Frequency and uptake of oral CF.

Table 1 and 2 summarize findings of seven descriptive studies of frequency and uptake of oral CF: Lyster and Ranta (1997); Ellis, Basturkmen, and Loewen (2001); Suzuki (2004); Sheen (2004); Lyster and Mori (2006); Panova and Lyster (2002); and Lee (2013). These studies represent diverse learner and setting backgrounds. An interesting pattern can be observed in Table 2. Namely, clarification requests, metalinguistic feedback, elicitation, and repetition lead to a greater rate of learner uptake than recasts, explicit correction, and translation. Common between these four oral CF strategies is that they prompt learners to correct their ill-formed utterances compared with the other three that directly supply learners with the correct form (Ferreira, 2006). In other words, answer-prompting oral CF strategies lead to greater uptake rates than answer-giving strategies because they are more explicit in requesting self-correction (Heift, 2004). Furthermore, implicit oral CF types such as recast may not be registered by learners as corrections of their ill-formed utterances and may be taken as discourse or meaning related (Allwright & Bailey, 1991; Calve', 1992; Chaudron, 1988; Lyster & Ranta, 1997; Netten, 1991).

Below is a brief discussion of important highlights about the frequency and uptake of seven oral CF strategies.

2.2.1.1. *Clarification requests.*

Clarification requests indicate to learners "that their utterance has been misunderstood by the teacher or that the utterance is ill-formed in some way and that a repetition or a reformulation is required" (Lyster & Ranta, 1997, p. 47). Clarification requests are implicit because they do not specify the source or location of the error. Examples of clarification requests include saying

“Pardon me?” or “Sorry?” in response to learner ill-formed utterances. Based on Table 1, clarification requests tend to occur with an intermediate frequency and have a high rate of uptake. This intermediate frequency could be because clarification requests fit naturally and easily in oral conversations. They are often used to indicate incomprehensibility in native-native communications and function as indirect requests for repetition. Furthermore, the high saliency of clarification requests could explain the high uptake rate. However, clarification requests could be misunderstood by learners to mean incomprehensibility due to mispronunciation, inattention of listener or other communication issues not related to formal errors in their utterances. This could lead to repetitions or reformulations that do not address the form-oriented purpose of clarification requests.

2.2.1.2. Elicitation.

Elicitation involves requesting correction of ill-formed utterances through repetition of learner utterance and pausing before the location of the error, asking learners about the correct form of an ill-formed utterance, or asking learners for a reformulation of their utterances (Lyster & Ranta, 1997). According to Table 1, elicitation has a positive correlation with proficiency level as such elicitation is used more frequently with high proficiency learners than low proficiency learners. In contrast, as shown in Table 2, elicitation leads to high uptake rate at all L2 proficiency levels.

Differences in frequency of elicitation between high and low proficiency levels could be attributed to two factors. First, more focus tends to be given to fluency over accuracy for low proficiency learners who are struggling in clearly expressing themselves. Second, low proficiency learners tend to make more formal errors than high proficiency learners. These two factors could lead to avoiding elicitation because of its potential to disrupt low proficiency

Table 1

Occurrence Frequency Rate of Different Strategies of Oral Corrective Feedback

Study		L&R (n=104)	E (n=24)	SU (n=31)	SH (n=10)	L&M (n=153)	P&L (n=25)	L (n=60)
Age / L2 Level		Teenage Mixed	Adult pre-/intermediate	Adult Intermediate	Adult Mixed	Teenage Mixed	Adult Low	Adult Advanced
Setting		FSL	ESL	ESL	EFL	FSL/JFL	ESL	ESL
Occurrence Frequency Rate	High Frequency	Recasts	Recasts	Recasts	Recasts	Recasts	Recasts	Recasts
		elicitation	explicit c.	clarification r.	explicit c.	elicitation, clarification r., metalinguistic, repetition	translation	explicit c.
		clarification r.	elicitation	elicitation	clarification r.	explicit c.	clarification r.	clarification r.
		metalinguistic	repetition	repetition	metalinguistic		metalinguistic	repetition
	Low Frequency	explicit c.	clarification r.	explicit c.	elicitation, repetition		elicitation	elicitation
		repetition	metalinguistic	metalinguistic			explicit c.	metalinguistic
							repetition	

L&R = Lyster & Ranta, 1997

E = Ellis, Basturkmen & Loewen, 2001

P&L = Panova & Lyster 2002

SU = Suzuki, 2004

L&M = Lyster & Mori, 2006

L = Lee, 2013

SH = Sheen, 2004

ESL = English Second Language

FSL = French Second Language

JFL = Japanese Foreign Language

learners from the more important task of expressing themselves. As learners become more capable of expressing themselves with ease, the frequency of elicitation increases because it has the advantage of pinpointing the location of the error or overtly requesting error correction. In turn, this overtness of elicitation would expectedly lead to high rate of uptake as shown in Table 2.

2.2.1.3. Explicit correction.

Explicit correction indicates that a learner's utterance is incorrect and provides the correct form (Lyster & Ranta, 1997). In Table 1, explicit correction occurs with a low frequency with young learners and leads to a low uptake rate. The low uptake rate may explain the low frequency of

Table 2

Uptake Frequency Rate of Different Strategies of Oral Corrective Feedback

Study		L&R (n=104)	E (n=24)	SU (n=31)	SH (n=10)	L&M (n=153)		P&L (n=25)	L (n=60)
Age / L2 Level		Teenage Mixed	Adult pre-/intermediate	Adult Intermediate	Adult Mixed	Teenage Mixed		Adult Low	Adult Advanced
Setting		FSL	ESL	ESL	EFL	FSL (n=104)	JFL (n=49)	ESL	ESL
Uptake Frequency Rate	High Frequency	elicitation	clarification r.	elicitation, clarification r., metalinguistic, explicit c., repetition	elicitation, clarification r., metalinguistic, repetition	elicitation, clarification r., metalinguistic, repetition	Recasts	elicitation, clarification r., repetition	elicitation, repetition
		clarification r.	metalinguistic, elicitation	Recasts	Recasts	Recasts	elicitation, clarification r., metalinguistic, repetition	metalinguistic	Recasts
	Low Frequency	metalinguistic	explicit c.		explicit c.	explicit c.	explicit c.	Recasts	clarification r.
		repetition	repetition					explicit c.	explicit c.
		explicit c.	Recasts					translation	metalinguistic
		Recasts							

L&R = Lyster & Ranta, 1997

E = Ellis, Basturkmen & Loewen, 2001

P&L = Panova & Lyster 2002

SU = Suzuki, 2004

L&M = Lyster & Mori, 2006

L = Lee, 2013

SH = Sheen, 2004

ESL = English Second Language

FSL = French Second Language

JFL = Japanese Foreign Language

explicit correction with young learners. L2 teachers of young learners may prefer to use CF strategies that lead to a higher rate of uptake. Furthermore, the low uptake rate of explicit correction could be resulting from the fact that explicit correction indicates that the listener understood the intended message and therefore signals to learners that no follow-up response is required. On the other hand, frequency and uptake rates of explicit correction significantly varies between studies with adult learners. These inconclusive results do not assist in identifying generalizable patterns of frequency and uptake of explicit corrections for this learner population.

2.2.1.4. Metalinguistic feedback.

Metalinguistic feedback provides information about the form of erroneous utterances with or without supplying the correct form (Lyster & Ranta, 1997). From Table 1, metalinguistic feedback occurs with a medium frequency and has a high rate of uptake except for advanced adult learners. Lee (2013) did not detect a single instance of metalinguistic feedback in advanced adult L2 classroom; therefore, the rate of uptake could not be determined in that setting. High uptake rate could be due to the overtness of metalinguistic feedback and that fact that corrections may not be provided with metalinguistic feedback pushing learners to self-correct. On the other hand, metalinguistic feedback may not be used with advanced L2 learners because they are expected to know relevant metalinguistic information about their errors.

2.2.1.5. Recasts.

Recasts involve restating corrected ill-formed utterances (Lyster & Ranta, 1997). All six studies unanimously agree that recast is the most frequent oral CF strategy used in language classrooms. In fact, recast constitutes about two-thirds of oral CF totals in most studies. The high frequency of recast could be attributed to its implicit, unobtrusive nature (Doughty, 2001; Doughty & Williams, 1998; Long, 1996). It enables instructors to provide positive evidence and correction of learner error in a single turn without interrupting meaning-oriented interactions. However, despite being the most frequently used oral CF strategy, recast has the lowest rate of learner uptake. Learner perception of recasts could explain the low rate of learner uptake. Learners could easily mistake recasts for non-corrective repetitions in meaning-oriented interactions. If learners perceive recasts as corrective, they are more likely to produce uptake in response (Egi, 2010) especially with adult and highly motivated learners (Suzuki, 2004).

2.2.1.6. Repetition.

Repetition involves restating ill-formed utterances, often accompanied by intonation change to highlight the error (Lyster and Ranta, 1997). Repetition tends to occur with low to medium frequency and leads to medium to high rates of uptake. Similar to recast, being implicit in nature, repetition may not be readily registered by learners as serving a corrective function. However, when stress and intonation are used to highlight form errors, repetition is rendered more explicit and could lead to greater uptake rate.

2.2.1.7. Translation.

The translation strategy of oral CF involves restating L1 learner utterances in L2 (Panova & Lyster, 2002). Translation occurs quite infrequently, with most studies not detecting a single instance of translation except for Panova and Lyster (2002) who found it occurs frequently in adult beginner classes. This suggests that L2 teachers are more likely to use translation with adult beginner L2 learners. Nevertheless, despite its frequent occurrence in Panova and Lyster, translation has a very low rate of uptake, meaning learners do not restate their L1 utterances in L2 in response to translation. Similar to recast, this low rate of uptake could be attributed to learner perception of translation as non-corrective repetitions.

The next section reviews studies that examined frequency and uptake of written CF.

2.2.2. Frequency and uptake of written CF.

Approximately half of written CF given by teachers focuses on form (Heffernan et al., 2014; Hyland, 2003). This rate reflects the importance of accuracy for L2 instructors. Two-thirds of written CF led to revisions in subsequent drafts and most of these revisions led to correct modifications (Hyland, 2003). However, unlike oral CF strategies, the frequency of written CF strategies has not been studied. The lack of interest in examining the frequency of written CF

strategies may be due to the asynchronous nature of writing, which allows ample opportunity for conscious planning and selection of written CF strategies and the fact that L2 writing instructors tend to use specific written CF strategies when evaluating their students' writings as opposed to freely alternating between different written CF strategies. For example, L2 instructors who identify errors by means of color coding tend to use this method consistently within and across their students' papers. Therefore, no value can be seen in studying the frequency of written CF strategies.

On the other hand, learner uptake following different written CF strategies has been examined by few studies (Sachs & Polio, 2007; Santos, Lopez-Serrano & Manchon, 2010; Storch & Wigglesworth, 2010). These studies have compared learner uptake following direct correction by means of underlining errors and providing correct forms, identification by means of editing codes, and reformulation by means of rewriting learners' sentences without errors. They found that direct correction and identification led to greater noticing and uptake by L2 learners than reformulation. The advantage of direct feedback and identification over reformulation may lie in the greater saliency and narrower focus of the former two strategies compared with the latter (Storch & Wigglesworth, 2010). Specifically, direct correction and identification provide feedback at the word level, but reformulation provides feedback at sentence or phrase level. No studies were found that compared learner uptake of other written CF strategies such as metalinguistic CF, which involves providing a technical explanation of the error and its type and concordance feedback, which involves presenting learners with concordance lines that show them correct use of the linguistic structure they are struggling with.

The following section reviews studies that compared the effectiveness of oral CF and written CF strategies in promoting L2 development. It presents a systematic review of the

literature through a meta-analysis of the combined statistical data reported by experimental and quasi-experimental studies.

2.3. Corrective Feedback and L2 Development: Meta-analysis

A meta-analysis approach is opted in for the literature review that examined the effectiveness of different oral and written CF strategies in promoting L2 acquisition. Systematic reviews have certain advantages over narrative reviews. First, systematic reviews are less prone to selective bias because the selection of studies is governed by clearly defined criteria. Well-articulated selection criteria ensure that studies are not excluded just because they do not support the main thesis of reviewers. Second, the combination of statistical data from multiple studies enables the detection of significant differences that otherwise could not be detected by individual primary studies due to the small sample sizes of individual studies. The combined sample size is larger in systematic reviews and thus provides more statistical precision. Third, systematic reviews can be used to detect and exclude outlier studies. The inclusion of outliers in a review is sometimes inadvisable due to the greater likelihood that an outlier suffers from sampling error.

In the last decade, six studies have systematically reviewed the literature of corrective feedback (Kang & Han, 2015; Li, 2010b; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006; Truscott, 2007). These reviews focused on written CF (Kang & Han, 2015; Truscott, 2007), oral CF (Li, 2010b; Lyster & Saito, 2010; Mackey & Goo, 2007), or both oral and written CF (Russell & Spada, 2006). They also examined some mediating factors such as learner age and language setting. All of these reviews, with the exception of Truscott (2007), have detected a medium to large effect size for corrective feedback indicating an effective role in L2 development.

The present review is set to provide an updated meta-analysis of CF literature since the last meta-analysis of oral CF was conducted five years ago and the most recent meta-analysis of written CF was limited to two types of written CF strategies.

The following sections are organized in accordance with the conventional organization of systematic reviews. First, a method section describes the process of identifying primary studies, the criteria for inclusion and exclusion of studies, and the procedures for coding studies and calculating effect sizes. Second, the results section reports findings according to language medium (written vs. oral) and CF strategies (direct, indirect, etc.).

In the present meta-analysis, oral and written CF and CF strategies are considered independent variables and L2 development as a dependent variable. The meta-analysis addresses the following questions:

- How effective is oral CF in promoting L2 development and accuracy?
- How effective are different oral CF strategies in promoting L2 development and accuracy?
- How effective is written CF in promoting L2 development and accuracy?
- How effective are different written CF strategies in promoting L2 development and accuracy?

2.3.1. Method.

2.3.1.1. Identification of primary studies.

Two methods were used to identify studies potentially suitable for inclusion into the present meta-analysis. Search was not limited to any posterior cut-off date. All primary CF studies published up until 2015 were included as a potentially suitable study for inclusion in the present meta-analysis. The goal was to be as comprehensive as possible in surveying available

literature. First, major academic publication repositories and peer-reviewed journals were searched to find potential studies using specific search keys and their combinations (see Appendix D. for a list of searched repositories and journals and used search keys). This search intentionally included unpublished Ph.D. dissertations as a countermeasure against ‘file drawer’ bias (also known as gray literature or publication bias) in which studies with insignificant statistical findings do not get published (Rosenthal, 1991). Second, the reference sections of key CF publications were scanned for potential primary studies.

2.3.1.2. Inclusion and exclusion criteria.

The set of inclusion and exclusion criteria used in the present meta-analysis has many items common with previous CF meta-analyses. Studies that met the following criteria were included in the present meta-analysis:

- It had at least one CF strategy as an independent variable.
- It measured L2 development in terms of grammatical accuracy or morphosyntactic acquisition. This inclusion criterion differs from Li’s (2010b) meta-analysis in which studies that measured L2 development in terms of lexical and phonological acquisition were also included.
- Its treatment involved meaning-oriented activity.
- Its treatment involved delivering CF orally or written in face-to-face or computer-mediated settings.
- It employed an experimental or a quasi-experimental design with pretests and posttests and a control or comparison group.
- It was published in English.
- It employed statistical analyses of mean differences.

- It reported the mean and standard deviation, gain scores or treatment effect size in Pearson's r or Cohen's d of experimental and control groups.
- It recruited child or adult L2 learners as participants in either second or foreign language settings.

Studies that met the following criteria were excluded from the present meta-analysis:

- Its treatment involved chat-based communication. Computer chatting is a unique medium of communication. It resembles oral conversation in being fast-paced and involving short turn-taking and resembles writings in visual manifestation. Thus, it was deemed unsuitable to include chat-based studies with either oral or written CF studies.
- For written CF studies, its treatment did not involve writing composition.

2.3.1.3. Coding.

Coding of primary studies was an iterative process in which initial coding helped in identifying key aspects that were included in subsequent iterations. Initially, coding focused only on types of CF strategies and language learning setting. Then, the process of initial coding led to identifying other features that could serve as moderator variables. Coding, in Lipsey and Wilson's (2001) terms, focused on substantive and methodological aspects. Substantive aspects tend to serve as independent variables while methodological aspects tend to serve as moderator variables. The coding protocol is summarized in Table 3, and two features used in the current meta-analysis are discussed in more detail below.

2.3.1.3.1. Corrective feedback strategies.

Identifying the type of CF strategies investigated in primary studies was not a straightforward task. This is mainly due to two reasons. First, not all studies use the same naming scheme for CF strategies. For instance, 'explicit and implicit CF' are used to describe recast and metalinguistic

Table 3

Meta-Analysis Coding Scheme

Features	Descriptors
CF strategy	Clarification request, metalinguistic feedback, recast, repetition, etc.
Language mode	oral vs. written
CF delivery mode	oral vs. written
Language learning setting	second language vs. foreign language
Participants' proficiency level	beginner, low-intermediate, high-intermediate, etc.
Participants' age group	child / teenager / adult
Target language	Arabic, Chinese, English, French, etc.
Target structure	articles, models, past tense, etc.
Communication medium	face-to-face vs. computer-mediated
Study design	experimental vs. quasi-experimental
Timing of posttests	immediate (first week after treatment), short-term (within 8-29 after treatment) and long-term (30 days or later after treatment)

feedback by some studies (e.g., Varnosfadrani & Basturkmen, 2009) while others refer to written direct corrections as 'red pen' technique (e.g., Maleki & Eslami, 2013). This naming issue requires examining samples of CF treatment if available to determine the exact type of CF strategies used. Second, some studies mix two or more CF strategies in the treatment of one experimental group (e.g., Lyster & Izquierdo, 2009). In this case, this combination of CF strategies was included in the meta-analysis if it was investigated by two or more studies.

2.3.1.3.2. Timing of posttests.

The variability of posttests' timings made it almost impossible to come up with all-inclusive descriptors. For instance, timings of first posttests ranged from immediately following the end of treatment to two weeks later. Thus, it was hard to draw an exact line where immediate posttests end and subsequent posttests start. However, in order to facilitate comparison with past CF meta-analyses, the present meta-analysis adopted Keck, Iberri-Shea, Tracy-Ventura, and Wambaleka (2006) posttests descriptors in which posttests held within the first week after treatment were considered as immediate posttests, within 8-29 days after treatment as short-term posttests, and 30 days or later after treatment as long-term posttests. Consequently, in the present meta-analysis, the mean and standard deviation of posttests held within the same timing category were averaged and used. For example, Mackey and Philp (1998) and Rahimi and Zhang (2013) had two delayed posttests held within the same timing category and therefore performance data on both tests were averaged and used as one posttest.

2.3.2. Analysis.

Comprehensive Meta-Analysis (CMA), a professional statistical software, was used for calculating the combined effect sizes in the present meta-analysis. CMA was developed by Borenstein et al. (2011) for the purpose of providing a comprehensive solution for conducting meta-analysis. CMA accepts over a hundred different types of data entry format, facilitating the process of meta-analyzing studies with varying statistical formats. Furthermore, it offers many useful utilities such as identifying publication bias by means of a funnel plot and performing Q-tests. This software has been widely used in many academic fields, including two past CF meta-analyses (Kang & Han, 2015; Li, 2010b). A key advantage of using CMA is that it minimizes the

likelihood of human errors creeping into the data analysis process especially when dealing with a large number of studies with different format types.

2.3.2.1. Meta-analysis models.

Fixed-effect model and random-effects model are two statistical models used in meta-analysis. These two models are based on two different assumptions about the effect size of individual studies included in the meta-analysis (Borenstein et al., 2011). On the one hand, the fixed-effect model assumes that all individual primary studies contribute the same effect size. This assumption could be true in cases where no or little variability exists across primary studies. On the other hand, random-effects model assumes that primary studies contribute different effect sizes due to variability in moderator variables such as population age and experimental settings. For the present meta-analysis, the second assumption is more likely to hold true as wide variability exists across the present sample of primary studies. Therefore, the random-effect model is opted in for the present meta-analysis.

2.3.2.2. Calculation of effect sizes.

The calculation of effect sizes depended on the format used for reporting statistical data by primary studies and the particular conditions of their experimental and control groups. First, four main reporting formats were used in the primary studies included in the present meta-analysis:

- mean scores and standard deviation of control and experimental groups on pretests and posttests measures.
- mean and standard deviation of gain scores between pretests and posttests of control and experimental groups.
- t value and F value.

Each format requires a specific type of mathematical equation for calculating effect size. Second, primary studies varied in the conditions of control and experimental groups. Some studies had the experimental groups receive some kind of additional instructional intervention beside CF. In this case, the group that received the same intervention minus CF was used as the control.

2.3.2.3. Sample size inflation.

A common problem encountered when conducting a meta-analysis is the inflation of sample size (Lipsey & Wilson, 2001). Sample size inflation refers to treating results from different moderating variables as belonging to separate samples when in fact they belong to a single sample. For instance, treating data from different outcome measures as data from separate samples leads to inflation of sample size. In order to avoid this problem, data from variables associated with a single construct were averaged to arrive at a single effect size before inclusion in the meta-analysis.

2.3.2.4. Outliers.

Outliers are studies contributing extreme effect values to the combined effect of the meta-analysis. Inclusion of such studies usually negatively impacts the results of meta-analysis especially in situations like CF studies where primary qualifying studies are few. Two steps were taken in order to identify outliers. First, normal probability tests were conducted on CF studies' data to determine if data are normally distributed. Once the normality of data was established, an outlier detection test was conducted using modified z-score. Modified z-score was opted for per Iglewicz and Hoaglin's (1993) recommendation as more suited for small sample sizes.

Modified z-scores were computed for subgroups in all studies in reference to posttest timings (immediate, short, or long) and language medium (oral vs. written). Per the

recommendation of Iglewicz and Hoaglin, modified z-scores above 3.5 were deemed outliers and excluded from analysis. The following formula was used for computing modified z-scores:

$$M_i = \frac{0.6745(x_i - \hat{x})}{MAD} \text{ (MAD is the median absolute deviation, } \hat{x} \text{ is the median)}$$

2.3.2.5. Publication bias.

Publication bias refers to the greater likelihood that studies with significant effect sizes get published (Dickersin, 2005). One way to mitigate the effect of publication bias is to include unpublished studies such as unpublished dissertations and conference presentations. The present meta-analysis included unpublished dissertations because they are publically more accessible. Additionally, CMA helps in visually identifying publication bias via funnel plots. A symmetric funnel plot indicates the absence of publication bias while fewer studies on the left side of the mean indicate an absence of studies with low effect sizes.

2.3.3. Results.

Results are grouped into two sections based on the medium of communication: oral or written. In each section, the combined effect sizes of CF at different posttest intervals are reported. Then, the effect size of CF strategies is reported and compared at different posttest intervals. To detect the presence of publication bias, funnel plots are presented and discussed for oral and written CF.

2.3.3.1. Oral CF and L2 development.

A total of 25 primary oral CF studies qualified for inclusion into the present meta-analysis. These studies are listed in Appendix E. and the CF strategies they examined are marked. The proportion of experimental and quasi-experimental studies were equal. Both experimental and quasi-experimental groups had pretests, posttests and control groups. However, they differed in whether participants were randomly sampled from target populations and

randomly assigned to experimental and control groups. Quasi-experimental studies often worked with intact classroom groups.

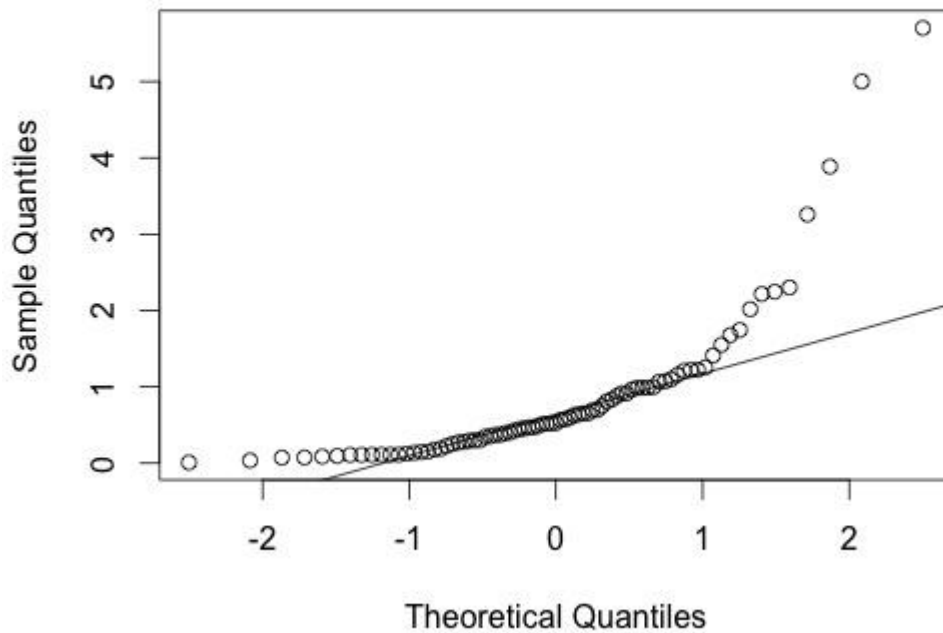


Figure 1. Normal probability Q-Q plot for oral CF studies

Outliers were identified using normal probability plot and modified z-score. The normal probability plot for oral studies (see Figure 1) shows an approximately normal distribution with a few extreme values. Modified z-scores were examined to identify outliers. Table 4 shows outlier subgroups that were excluded from analysis. In order to retain as much data as possible due to small sample size, outliers excluded from overall effect size analysis were included in the effect size analysis of individual subgroups after passing through a second outliers detection test that compared them only with studies within the same subgroup. This way studies conforming to their class of subgroups were retained.

Publication bias was tested via a funnel plot. A funnel plot for oral CF studies is presented in Figure 2. Studies with low effect size are located on the left side of the plot while those reporting large effect size are located on the opposite side. Publication bias exists when the

Table 4

Outlier Oral CF Studies

Posttest timing	Subgroup	Study
Immediate	ML	Li, 2013
		Rassaei & Moinzadeh, 2011
		Rassaei, 2014
	DC	Rassaei, 2013
		Yilmaz, 2013
Short	PRM	Rahimi & Zhang, 2013
Long	RC	Yilmaz, 2013

left side of the plot has fewer studies than the right side. This funnel plot appears symmetric with studies equally distributed around the mean, which is evident of lack of publication bias.

However, studies are clustered at the top of the plot, which means this collection lacks studies with smaller sample size.

Cohen's d effect size falls between three main categories: small ($d = .2$), medium ($d = .5$) and large ($d = .8$) (Cohen, 1988). The present meta-analysis returned a moderate overall effect size for oral CF on immediate ($d = 0.584$), short-term ($d = 0.671$) and long-term ($d = 0.576$) posttests. Table 5 shows Cohen's d values for oral CF across three posttests intervals along with the number of groups (k) included in the analysis, standard error values, and the confidence interval range.

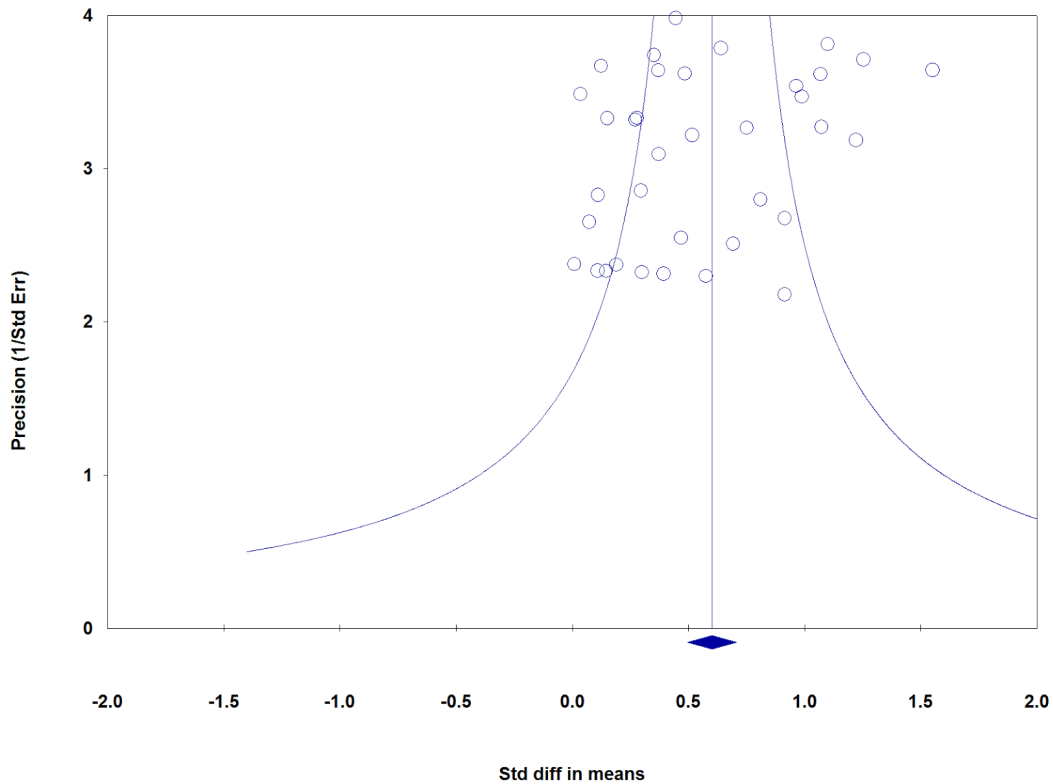


Figure 2. Funnel plot of precision by standard difference in means of oral CF studies

2.3.3.1.1. Effect of oral corrective feedback.

The effect size of five oral CF strategies was examined in the present analysis: clarification requests, direct corrections, metalinguistic feedback, prompts (i.e., a group of CF strategies that prompts learners to attend to form), and recasts. Repetition was excluded because it was only examined by a single study. Most of the included studies have examined metalinguistic feedback and recasts yielding greater precision for these two strategies and more confidence in their combined effect sizes. Direct correction, on the other hand, has been examined by only two studies and therefore has a larger margin of error.

The effect of different oral CF strategies varied as shown in Table 5. Overall, Table 5 shows a moderate to large effect for oral CF strategies. Direct corrections had a large immediate effect

size. Metalinguistic feedback had a large immediate and short-term effect size and a small long-term effect size. Clarification requests and prompts had moderate effect size across different

Table 5

Effects of Oral Corrective Feedback

Independent variable	k	Mean d	SE	95% Confidence interval	
				Lower limit	Upper limit
Overall Effect					
Immediate	36	0.584	0.071	0.444	0.723
Short	26	0.671	0.099	0.478	0.864
Long	12	0.576	0.143	0.296	0.856
Clarification requests					
Immediate	3	0.486	0.162	0.170	0.803
Short	3	0.514	0.170	0.181	0.848
Long	2	0.694	0.214	0.275	1.113
Direct Corrections					
Immediate	2	4.387	1.218	1.999	6.775
Metalinguistic feedback					
Immediate	11	1.055	0.258	0.549	1.561
Short	7	0.856	0.222	0.421	1.292
Long	3	0.308	0.188	-0.059	0.676
Prompts					
Immediate	4	0.692	0.182	0.335	1.049
Short	2	0.663	0.263	0.148	1.177
Recasts					
Immediate	21	0.557	0.098	0.366	0.749
Short	13	0.608	0.156	0.302	0.914
Long	5	0.396	0.183	0.037	0.755

posttest timings. Finally, recasts had a moderate immediate and short-term effect size and a small long-term effect size.

2.3.3.2. Written CF and L2 development.

A total of 16 primary written CF studies were included in the present meta-analysis (4 experimental and 12 quasi-experimental). Similar to oral CF studies, publication bias and outliers detection tests were conducted for written CF studies. The results of the two tests are reported next.

Normality assumption was validated by constructing a normal probability plot. The plot shown in Figure 3 provides a visual representation of the data and helps in preliminary identification of outliers. The plot validates the normality assumption of the present data and shows the existence of a number of potential outliers. Modified z-scores identified the short-term posttest data of the direct correction subgroup in Ellis et al. (2008) as an outlier. This subgroup was excluded from subsequent analysis.

Publication bias was tested through a funnel plot of included studies. The funnel plot shown in Figure 3 is relatively symmetric in the distribution of studies around the mean axle. This indicates that the studies included here are somewhat equally distributed between those reporting smaller effects and those reporting larger effects. However, missing from the data are studies representing smaller sample sizes.

2.3.3.2.1. Effect of written corrective feedback.

Meta-analysis results show a moderate to large effect for immediate ($d = 0.747$), short-term ($d = 0.835$), and long-term ($d = 0.733$) written CF on morphosyntactic L2 development and L2 writing accuracy. However, written CF strategies show considerable variation in their effect sizes. Indirect correction, direct correction, and metalinguistic feedback had small, moderate and

large effect sizes across all posttest timings, respectively. Most studies have examined direct corrections ($n = 15$) and metalinguistic feedback ($n = 11$) and fewer studies have examined indirect corrections ($n = 5$).

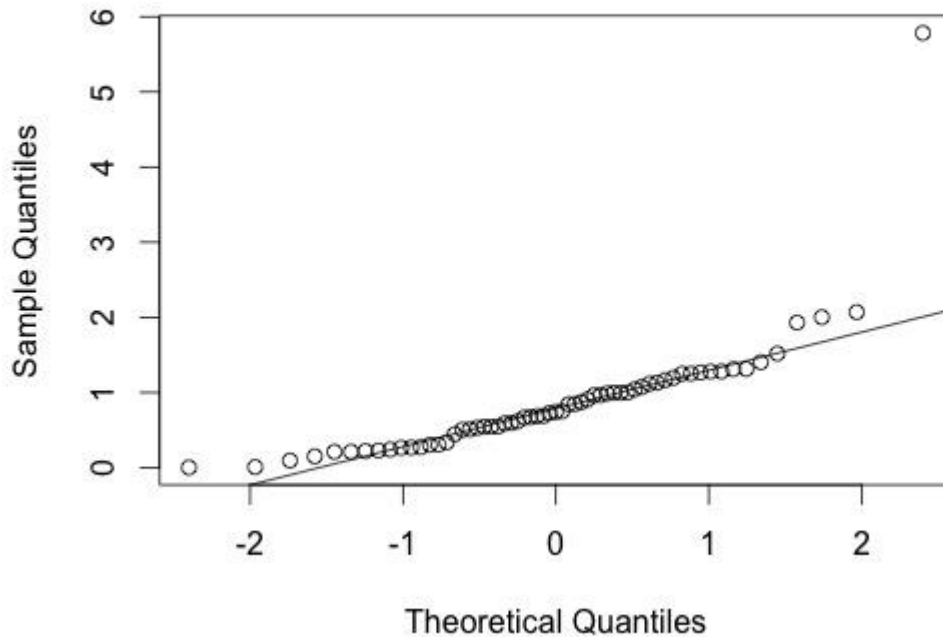


Figure 3. Normal probability Q-Q plot for written CF studies

2.3.4. Discussion.

This section discusses the results of oral and written CF reported earlier as they relate to the findings of similar past meta-analyses.

2.3.4.1. Oral CF discussion.

The first question posed by this review asked about the effectiveness of oral CF in promoting L2 development and accuracy. The present meta-analysis aligned with some past CF meta-analyses (Li, 2010b; Lyster & Saito, 2010; Russell & Spada, 2006) in identifying a durable moderate overall effect for oral CF on morphosyntactic development and accuracy. The findings of this and similar CF meta-analyses stand in sharp contrast with Truscott's (2007), which reported a surprisingly negative effect for oral CF. However, mounting evidence for the

effectiveness of oral CF in promoting L2 development overshadows Truscott's (2007) findings.

The second question asked about how oral CF strategies compare to each other in their effect on

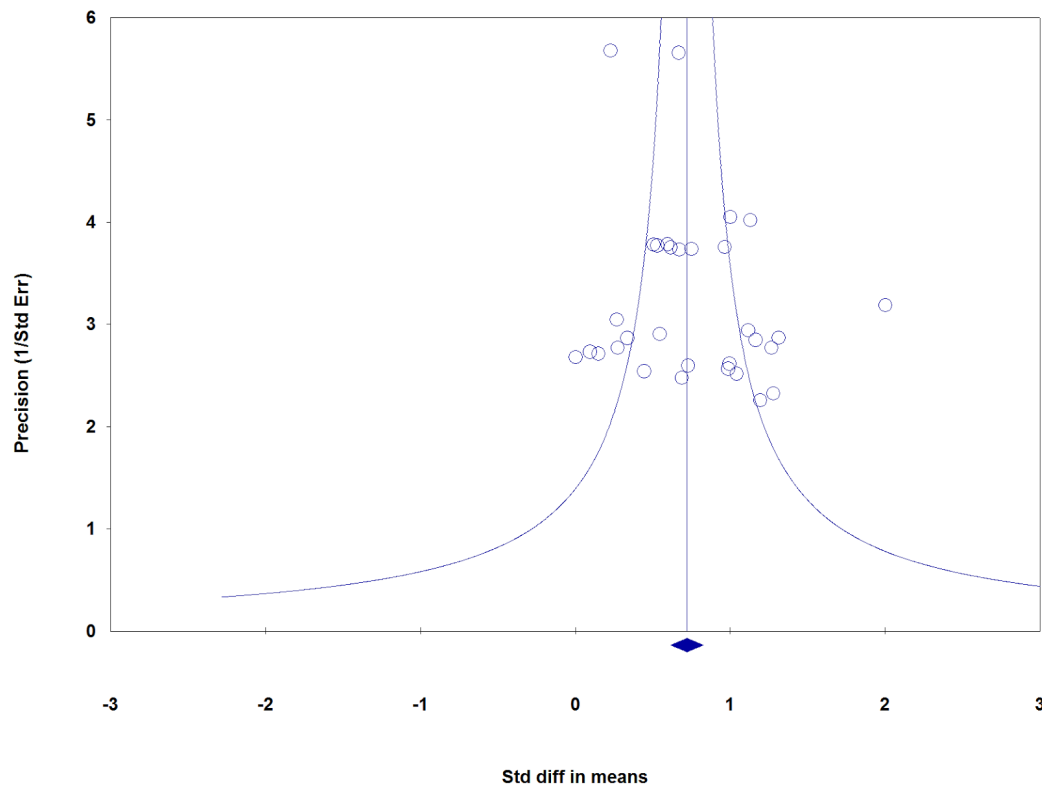


Figure 4. Funnel plot of precision by standard difference in means of written CF studies

L2 morphosyntactic development and accuracy. The present meta-analysis identified considerable differences in the effect of different CF strategies. As shown in Table 6, direct

Table 6

Effects of Written Corrective Feedback

Independent variable	k	Mean d	SE	95% Confidence interval	
				Lower limit	Upper limit
Overall Effect					
Immediate	31	0.747	0.078	0.595	0.899
Short	19	0.835	0.109	0.620	1.049
Long	10	0.733	0.158	0.423	1.043

Table 6 continued**Direct Corrections**

Immediate	15	0.776	0.118	0.545	1.007
Short	9	0.749	0.167	0.422	1.076
Long	4	0.694	0.246	0.212	1.175

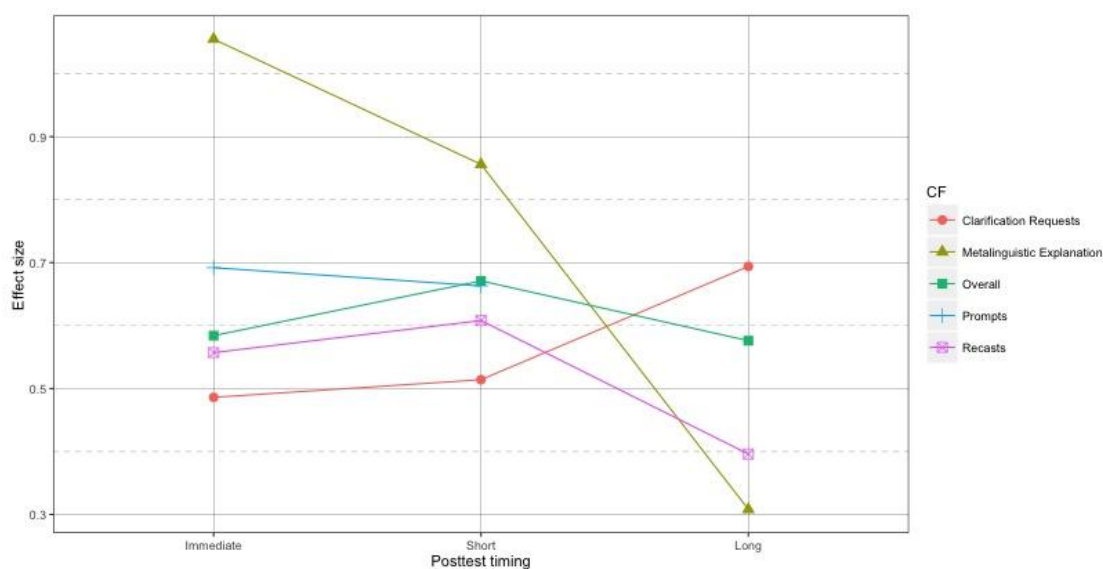
Indirect Corrections

Immediate	5	0.323	0.126	0.076	0.570
Short	3	0.335	0.145	0.051	0.619

Metalinguistic feedback

Immediate	11	0.880	0.106	0.672	1.088
Short	7	1.161	0.119	0.929	1.394
Long	5	0.886	0.252	0.392	1.381

correction and metalinguistic feedback had a large immediate effect. This significant effect of these two CF strategies could be attributed to their greater explicitness and saliency (Carroll & Swain, 1993; Ellis et al., 2006). As such, these two CF strategies leave no room for learners to misinterpret the intent of CF as meaning-oriented like in the case with implicit CF strategies such as recasts and clarification requests.

*Figure 5. Effects of oral CF strategies over time*

Another interesting pattern for oral CF strategies can be observed in the diagram of effect sizes for oral CF strategies shown in Figure 5. In the diagram, metalinguistic feedback exhibited a drastic decline in effect size between immediate and long-term posttests while recasts showed a moderate decline and clarification requests showed a moderate increase in effect size. This drastic decline may indicate that metalinguistic feedback promotes conscious awareness of morphosyntactic rules but not necessarily the internalization of these rules by learners. As a consequence, decline in conscious morphosyntactic knowledge leads to decline in the performance of L2 learners. This pattern has not been observed by past CF meta-analyses because none had enough studies that examined the long-term effect of metalinguistic feedback. In the present meta-analysis, only three primary studies contributed to the long-term combined effect size for metalinguistic feedback. Obviously, more long-term studies are needed for confidently determining the long-term effect of metalinguistic feedback and other oral CF strategies.

2.3.4.2. Written CF discussion.

The third question in this review asked about the effectiveness of written CF in promoting L2 development and accuracy. In line with Kang and Han (2015) and Russell and Spada (2006), results show a moderate to large overall effect for written CF. This effect size shows that written CF is very effective in promoting L2 morphosyntactic development and writing accuracy. Furthermore, the effect was durable suggesting that written CF led to internalizing of L2 morphosyntactic rules by learners. This underscores the importance of CF in promoting L2 writing accuracy.

The fourth question asked about the effectiveness of written CF strategies in promoting L2 development and writing accuracy. The diagram in Figure 6 shows differences in the effect

sizes of written CF strategies. Respectively, direct correction and metalinguistic feedback had a durable moderate and high effect sizes across all three posttests timings. Indirect corrections had a durable moderate effect size on immediate and short-term posttests. However, the long-term effect size of indirect corrections was not computed here due to an insufficient number of primary studies that explored its long-term effect.

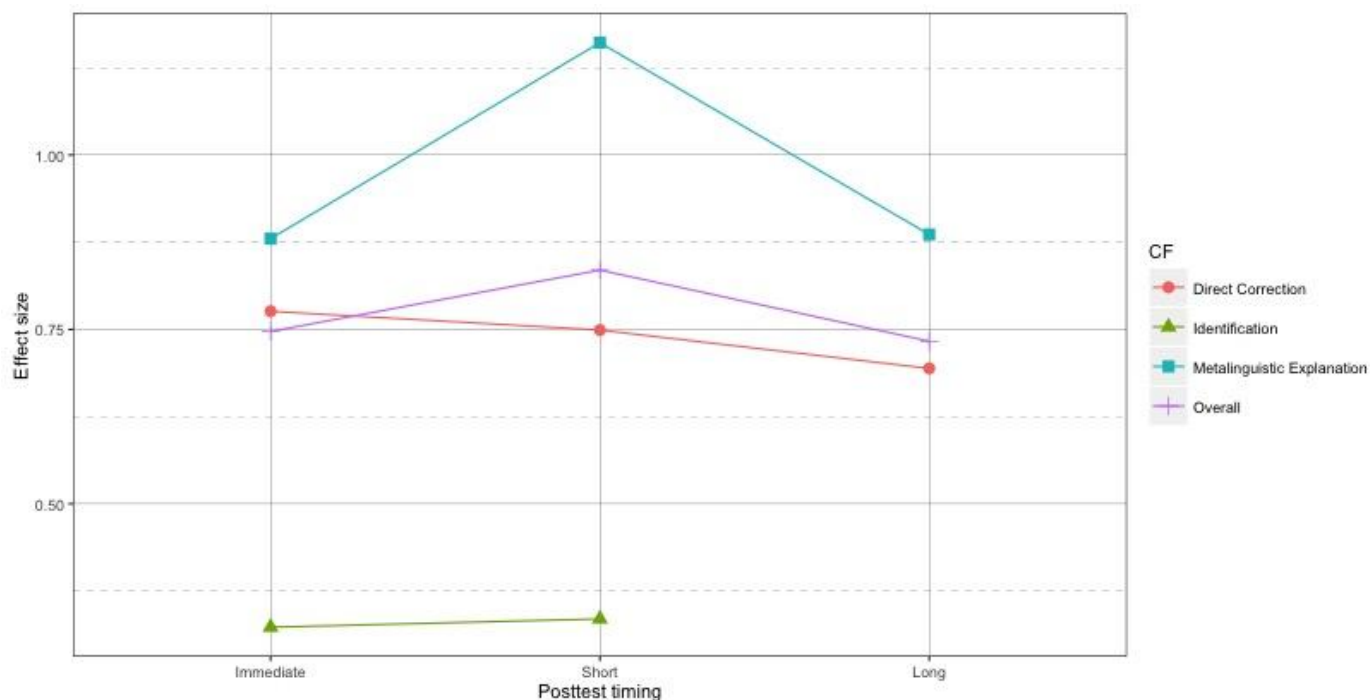


Figure 6. Effects of written CF strategies over time

Unlike oral CF strategies, all written CF strategies show a long-term persistent effect on L2 development. The persistent effect of written CF strategies could be attributed to the greater saliency of form and lesser time pressure on production in writing compared with speech. As such the writing medium gives learners more room to consciously reflect on their language production and apply rules learned from written CF. Moreover, Figure 6 shows that metalinguistic feedback had a substantially greater effect size than direct and indirect

corrections. This difference could be due to the fact that metalinguistic feedback promotes L2 learners' metalinguistic awareness of morphosyntactic rules and therefore give them the facility to evaluate the well-formedness of their L2 writings.

2.3.5. Conclusion.

The systematic review conducted here intended to determine the effect of oral and written CF and their strategies on L2 morphosyntactic development. For this purpose, it meta-analyzed 24 oral CF studies and 16 written CF studies. Meta-analysis results identified a positive moderate to large effect for both CF types. It also found a large durable effect for written metalinguistic feedback on L2 writing accuracy but a non-durable effect for oral metalinguistic feedback on the accuracy of oral L2 production.

The section above reviewed studies that examined the effect of CF and CF strategies on L2 morphosyntactic development. It identified positive effect for CF and variable effects for different CF strategies. The following section reviews studies that investigated the effectiveness of CF and CF strategies in Computer-Assisted Language Learning (CALL) settings.

2.4. CF in CALL

A survey of the literature identified a total of 17 studies that have investigated the effectiveness of CF in promoting morphosyntactic L2 development in CALL settings. However, close examination identified many methodological issues that undermine the findings of most of these studies. These issues range from lack of control groups (Nagata, 1993; Vaezi & Abbaspour, 2015; Trofimovich, Ammar & Gatbonton, 2007), control group receiving CF (Ferreira, 2006), use of non-meaning-oriented treatment tasks (Rosa & Leow, 2004; Sanz & Morgan-Short, 2004; Kregar, 2011; Sagarra, 2007), inadequate reporting of treatment task (Ferreira 2006), unnatural laboratory-based treatment tasks (Trofimovich, Ammar & Gatbonton,

2007), treatment including explicit grammar instruction (Hartshorn et al., 2010), inadequate isolation of experimental conditions (Vlugter et al., 2009) and lack of pretests (Yilmaz, 2012). Exclusion of methodologically problematic studies left a total of 7 studies for this review section (AbuSeileek, 2013; AbuSeileek & Abualsha'r, 2014; Loewen & Erlam, 2006; Monteiro, 2014; Penning de Vries, 2015; Sauro, 2009; Vlugter et al., 2009).

The effectiveness of two types of CALL-based CF has been examined by the six studies reviewed here: computer-mediated CF and computer-generated CF. Computer-mediated CF involves receiving CF from L2 instructors or peers while or after completing a language activity via a video-conference (Monteiro, 2014), computer-chat (Loewen & Erlam, 2006; Sauro, 2009) or corrected drafts (AbuSeileek, 2013; AbuSeileek & Abualsha'r, 2014). Computer-generated CF involves automated computer analysis of learners' speech (Penning de Vries, 2015) or writings (AbuSeileek, 2013) and provision of appropriate CF. The findings of these studies are reviewed below according to their language learning settings.

Loewen and Erlam (2006) and Sauro (2009) investigated the effect of receiving recasts and metalinguistic explanation for erroneous utterances during computer-chat. They found no significant gains from receiving computer-mediated CF and no significant differences in L2 development between recasts and metalinguistic explanation experimental groups. Loewen and Erlam (2006) suggest that some factors could have mediated the effect of CF, including learners being developmentally unready to acquire target structures, learners lacking sufficient pre-treatment knowledge about target structures to benefit from CF, and "reduced immediacy" (p. 10) of CF. Future research needs to explore how these and other factors mediate the effectiveness of CF in computer-chat settings.

Monteiro (2014) investigated the effect of recasts and metalinguistic explanation in video-conferencing language activities in promoting L2 syntactic development. She recruited 42 low-intermediate adult EFL speakers. Treatment tasks consisted of communicative jigsaw language activities. She found no statistically significant differences between treatment and control groups. Monteiro suggests that the form-focused tasks used in treatment may have contributed to the development of target structure in control group to a degree that rendered the contribution of CF conditions insignificant. Further research is needed to determine if indeed form-focused language tasks can have an impact equivalent to CF in video-conferencing settings.

In writing settings, a significant positive effect was found for both computer-mediated and computer-generated CF on the development of L2 writing accuracy (AbuSeileek, 2013; AbuSeileek & Abualsha'r, 2014). Treatment tasks in both studies involved having participants write short essays (150–200 words) and having these essays reviewed and feedback provided by peers. AbuSeileek (2013) investigated the effect of three CF treatments on L2 writing accuracy: reformulation via track changes, metalinguistic explanation with direct correction provided by a word processor, and a combination of these two. He recruited 64 EFL learners with no specific proficiency level. He found that receiving a combination of reformulation and metalinguistic explanation led to greater accuracy development than each CF treatment alone. He also found that all CF treatment types led to significant accuracy development compared with the control group. On the other hand, AbuSeileek and Abualsha'r (2014) investigated the effect of reformulation via track changes, reformulation without track changes, and metalinguistic labels of error types on L2 writing accuracy. All CF treatments led to significant accuracy development compared with control group. However, they reported greater effect for reformulation with track changes than reformulation without track changes followed by metalinguistic labels.

Penning de Vries (2015) investigated the effect of prompts by a computer tutor on L2 morphosyntactic development during automated spoken question-and-answer dialog. Treatment task involved a computer tutor asking participants about the content of a video clip and participants speaking their answering to a computer recorder. If the automatically recognized and parsed responses are grammatically incorrect, participants are prompted for correction. Penning de Vries found the effect of prompts in automated dialogs is mediated by L2 proficiency level. Namely, learners with low and intermediate proficiency level were found to benefit significantly from automated prompts more than learners with advanced L2 proficiency level.

This section reviewed studies that examined the effectiveness of both computer-mediated and automated CF and its different strategies in promoting morphosyntactic development. The next section reviews the theoretical basis of graduated CF in both traditional and automated settings. Then, it reviews studies that empirically investigated the effectiveness of graduated CF in promoting L2 development.

2.5. Use of Sociocultural Theory in CALL and CF

This section starts by defining Vygotsky's Sociocultural Theory (SCT) and discussing how its principles are applied to L2 research. Then, it discusses how SCT and its principles are used in CALL and CF research. This discussion segues into a review of graduated CF research. The main purpose of this section is to present the theoretical base and empirical evidence motivating the introduction of graduated CF in AWE tools, which is one of the main objectives of the present study.

2.5.1. Overview of Sociocultural Theory.

SCT argues that social environments and communication tools mediate the construction of meaning (Vygotsky, 1987). The main premise of SCT is that meaning construction does not

take place in a vacuum but in fact it is directly shaped by the social and cultural dimensions of human activities and the tools used for carrying out such activities. In other words, SCT makes no separation between mental activities and their social and cultural contexts. This perspective views learning as participation (Donato, 2000). Increased and successful participation within a community of practice is considered evidence of learning. Three principles of SCT are used for understanding the process of meaning construction and learning: mediation, social learning, and genetic analysis (Warschauer, 2005). These principles are examined below in light of their applicability to CALL and CF studies.

2.5.1.1. Mediation.

Mediation is “the process through which humans deploy culturally constructed artifacts, concepts, and activities to regulate (i.e., gain voluntary control over and transform) the material world or their own and each other’s social and mental activity” (Lantolf & Thorne, 2006, p. 79). This means that the significance of socially and culturally constructed tools and artifacts does not lie in and of the tools and artifacts themselves, but lies in the forms of activities they mediate and bring about. It is argued that the mediational quality of sociocultural artifacts with the help of competent language users socializes novice language users into a community of practice and successful socialization and participation is a sign of learning (Donato & McCormick, 1994). To illustrate this notion, take the example of smartboards versus traditional blackboards.

Smartboards have transformed the activity of presenting information in the classroom into a more interactive and multimodal activity that allows for newer forms of meaning construction. Because a smartboard, in this example, is seen as a sociocultural artifact mediating classroom activity, its use and perception is believed to be significantly shaped by social and cultural dimensions.

2.5.1.2. Social learning.

SCT argues that learning is a social phenomenon that first takes place between individuals prior to being internalized by novice learners (Vygotsky, 1978). This process of development involves scaffolding and adjusting the activity to learners' Zone of Proximal Development (ZPD). ZPD refers to the difference between what a learner can perform alone and what a learner can perform with help from others. At the appropriate ZPD level, it is argued that a novice learner is capable of performing new skills and acquiring new knowledge not achievable otherwise. This process of internalization, known as appropriation in the sociocultural literature, requires dialogic interaction between expert users and novice learners (Aljaafreh & Lantolf, 1994; Newman, Griffin & Cole, 1989).

2.5.1.3. Genetic analysis.

SCT argues that the construction of meaning can only be fully understood by analyzing the historical, cultural, and developmental dimensions of the activity in which meaning is constructed (Kozulin, 1991; Warschauer, 2005). Without careful analysis of these dimensions, many aspects of meaning construction will be difficult to interpret. Examples of these dimensions could include such historical and cultural events as the introduction of overhead projectors in public school classrooms and their subsequent use by instructors and students in presenting information. By reflecting on the historical and cultural dimensions of using overhead projectors in the workplace, one can understand the motivation, or the lack thereof, of using overhead projectors by instructors and students in public schools.

2.5.2. Use of Sociocultural Theory in CALL.

SCT brings a unique perspective to the study of CALL. From a SCT perspective, CALL is not basically language learning plus technology, but it is in fact a new form of sociocultural

activity that warrants careful consideration of its social, cultural, and developmental dimensions (Warschauer, 2005). The following will discuss how the three principles of SCT are applied to CALL research.

2.5.2.1. Mediation and CALL.

From the mediation perspective, technology is not viewed as just a means of enhancing traditional forms of meaning construction but as a means of mediating newer forms and a means of socializing novice users into a community of practice (Warschauer, 2005). In other words, the use of technology transforms the flow and nature of traditional L2 teaching and learning activities. For example, the use of Computer-Mediated Communication (CMC) mediates the development of a type of L2 writing literacy uniquely different from the type mediated by traditional paper-based literacy activities (Shetzer & Warschauer, 2000). Successful participation and meaning construction in CMC entail appropriate use of information technology and sociocultural norms.

2.5.2.2. Social learning in CALL.

The notion of social learning has been extensively used in studying Second Language development as it enables researchers to determine how expert users such as L2 teachers and competent L2 speakers adjust input to the ZPD level of novice learners and how this process leads to L2 development (Aljaafreh & Lantolf, 1994). For CALL, the notion of social learning has been extensively used in studying Computer-Mediated Communication (CMC) such as online forum discussion, chatting, and video-conferencing (Warschauer, 2005). This notion helps in understanding how CMC interactions with competent language users lead novice language users to incorporate and refine new linguistic knowledge and skills.

A look at past CALL studies employing the notion of social learning shows that focus of research was primarily on CMC settings. No studies, to this author's knowledge, have explored L2 learning in Human-Computer Dialog (HCD) settings using the notion of social learning. Employing the notion of social learning in analyzing language learning in HCD settings offers the advantage of exploring the social dimension of learning in those settings. Further, it helps unravel the process of learning by linking learning with observable social human activities. But, most importantly, the notion of social learning goes beyond other language learning theories such as the Interaction Hypothesis in emphasizing the mediational role of social human interactions and its effect on L2 learning. In other words, while the cognitive-interactionist perspective emphasizes the importance of the way content and knowledge is delivered to learners, the sociocultural perspective emphasizes the role of the social dimension of that process (Erlam et al., 2013).

CALL studies that investigated the effectiveness of collaborative learning in the ZPD found encouraging results. Belz and Kinginger (2002) investigated the development of German address form use by L1 English speakers in a CMC context. They found that peer assistance in the form of CF led to observable improvement in the use of German address form. Further, Belz and Kinginger pointed that learners' interactions with competent language speakers in CMC context exhibited legitimate peripheral participation. Legitimate peripheral participation refers to the gradual movement of new members (i.e., L2 learners) from peripheral to central participation as a consequence of engagement in a social activity and observing accepted patterns of language use (Lave & Wenger, 1991). Many other studies added supporting evidence to the positive effect of L2 learners' interactions in CMC settings on L2 learning (e.g., Cuestas Verjano, 2013; Peterson, 2012; Zeng & Takatsuka, 2009). These studies were conducted in different types of

CMC, including synchronous and asynchronous settings and text-based and video-based modalities.

2.5.2.3. Genetic analysis in CALL.

Genetic analysis is used in L2 research to understand and determine the place and importance of sociocultural artifacts in light of their historical, cultural, and developmental backgrounds. For instance, genetic analysis helps in determining what motivates the choice between different language styles, language teaching approaches, and types of language interactions. In CALL research, genetic analysis is used to determine and understand factors affecting the motivation behind technology use in L2 teaching and learning (e.g., Warschauer, 1996). Further, genetic analysis is used to understand how identity is constructed and maintained in CALL settings and how agency is negotiated (Kern & Warschauer, 2000). In an interesting study that illustrates the use of genetic analysis in CALL, Thorne (2003) discovered that, despite being a ubiquitous medium of communication, email has failed to mediate age peer relationships in his study due to a “generational shift in communication tool preference” (p. 38).

2.5.3. Use of Sociocultural Theory in CF.

SCT brings a unique perspective to CF research by viewing correction as an activity carried out with learners as opposed to an activity done to learners (Erlam et al., 2013). In other words, during error correction both competent users and novice learners are actively and mutually engaged in fulfilling the objective of error correction, namely, adjusting CF to learners' ZPD, and promoting self-correction (Aljaafreh & Lantolf, 1994). This SCT perspective on CF sharply contrasts with the cognitive-interactionist perspective, which views CF as a number of unique strategies that have a variable effect on L2 learning rigidly applied without consideration to the fluid nature of social human interactions (Lyster & Ranta, 1997). The SCT perspective

clearly offers the advantage of tapping into the social dimension of human interactions to maximize learning from CF.

The three principles of SCT outlined earlier can be used for analyzing the sociocultural dimensions of CF. First, CF can be examined for mediational qualities that socialize L2 learners into their respective target community of practice. In other words, the notion of mediation can explain how CF from competent language users mediates positive change in the use of linguistic resources by novice language users that consolidate their membership in the target community of practice. Second, CF activities can be analyzed as a social learning activity where CF is adjusted as dialogic interactions unfold to reach the ZPD levels of novice learners. Further, the notion of social learning helps to determine if CF leads to learning as measured by increased participation in the community of practice and decreased dependence on the help of competent language users (Erlam et al., 2013). Third, genetic analysis can be employed to determine the historical and cultural backgrounds affecting the perception and use of CF. For instance, genetic analysis can uncover if the decision of L2 teachers to provide CF to L2 learners and the manner of providing CF are related to such historical and cultural events and attitudes as the introduction and subsequent wide adoption of communicative and natural language teaching approaches that emphasize fluency over accuracy.

Clearly, CF research can benefit from adopting a SCT perspective. Unfortunately, CF studies that adopted a SCT perspective are very sparse. The following section reviews the findings of these studies.

2.5.3.1. Graduated CF.

The sociocultural perspective to effective intervention suggests that effective intervention should have three characteristics (Aljaafreh & Lantolf, 1994). First, effective intervention should

be graduated and attuned to learners' needs with the purpose of providing the minimal guidance needed for task completion. Second, effective intervention should be contingent on learners' need as such intervention should be retracted once learners show evidence of self-reliance in overcoming their own language difficulties. Third, effective intervention should be dialogic in order to identify learners' Zone of Proximal Development (ZPD) and thus successfully provide graduated and contingent feedback. This sociocultural approach to intervention applied to CF, which came to be known as graduated CF, produced motivating results and showed greater advantage in promoting L2 development than non-graduated CF (Erlam et al., 2013; Han, 2001; Muranoi, 2000; Nassaji & Swain, 2000; Rassaei, 2014).

The three characteristics of effective intervention outlined above can be seen in two educational contexts: student-teacher writing conferences and dialog-based language activities. In both writing conferences and dialog-based language activities, teachers are able to fine-tune feedback to meet the proficiency levels and needs of their students by starting with the least explicit feedback strategy, questioning students about what they know about their errors, and leading them to self-correct. In writing conferences, a number of studies had learners receive graduated CF on their grammatically incorrect sentences to determine its effect in promoting L2 development in comparison with non-graduated CF (Aljaafreh & Lantolf, 1994; Erlam et al., 2013; Han, 2001; Nassaji & Swain, 2000). Other studies had learners receive graduated CF during dialog-based activities (Muranoi, 2000; Rassaei, 2014). In both contexts, graduated CF was determined to lead to greater L2 morphosyntactic development and to be more effective in promoting self-correction than non-graduated CF.

All in all, graduated CF is a dynamic and personalized approach for delivering traditional strategies of CF such as recasts and metalinguistic feedback. It seeks to deliver level-appropriate

feedback when needed with an eye toward promoting self-reliance. However, this advantage comes with an important caveat. Graduated CF requires one-to-one dialog with learners, which means more time and effort by L2 teachers. It may be feasible to provide graduated CF in situations where the teacher-to-student ratio is low. But, oftentimes, this is not the case. In such situations, a good solution would be to provide graduated CF via a computer tutor. In addition to solving the logistical difficulty of delivering graduated CF to a large number of L2 learners, interaction with computer tutors may prove less face-threatening especially with introvert learners.

A survey of the literature failed to identify any attempt to explore how automated graduated CF would be received by L2 learners and what factors shape their perceptions of graduated CF. Clearly, this is a neglected but important area of investigation that has substantial implications in CALL contexts. For this reason, the present study developed and piloted automated graduated CF in order to gauge L2 learners' interest in graduated CF and explore ways to improve it.

Graduated CF leads to the bigger topic of preferences and perceptions of automated CF and CF strategies. Interesting is how L2 learners' preferences and perceptions of graduated CF compare to other CF strategies and what concerns and factors influence their preferences and perceptions. The next section reviews the findings of studies that examined L2 learners' preferences and perceptions of CF and different CF strategies. Findings of perceptions and preferences studies are of direct relevance to the topic of this dissertation.

2.6. Perceptions and preferences of CF and its strategies

L2 researchers tend to agree that perceptions, attitudes, and beliefs of L2 learners about what constitutes effective language learning and teaching practices have a significant effect on

learning outcomes (Amrhein & Nassaji, 2010; Brown, 2009; Schulz, 2001). If these affective factors are not met or addressed, learning can be significantly impaired due to learners' frustration, demotivation, anxiety, or distrust in the learning or teaching process (Green, 1993; Horwitz, 1990; Kern, 1995; McCargar, 1993; Noel, 2001; Noels, Clément, & Pelletier, 1999; Tse, 2000).

CF is an example of a common L2 teaching and learning practice that may be influenced by learners' perceptions of its effectiveness and their preferences of its different strategies and qualities. The effectiveness of CF can be negatively impacted if a mismatch exists between what L2 learners' perceive as an effective and preferred CF quality and strategy and the kind of CF quality and strategy that they actually receive (Amrhein & Nassaji, 2010). L2 learners are less likely to pay attention or use CF if they give little value to it and the opposite is true (Amrhein & Nassaji, 2010; Kartchava & Ammar, 2014a).

Surprisingly, research has shown that mismatches in fact widely exist between L2 teachers and learners about what they consider as desirable and effective CF qualities and practices (Agudo, 2012; Diab, 2005; Jeon & Kang, 2005; Park, 2010; Plonsky & Mills, 2006). For instance, while L2 learners tend to prefer more explicit and direct CF, their L2 instructors tend to value the opposite types of CF (Jeon & Kang, 2005). This mismatch needs to be addressed in order to render CF more effective. It can be addressed by either providing L2 learners with the kind of CF quality and strategy that they prefer and perceive more effective or changing their attitudes towards CF to align with research findings about effective CF qualities and practices. Indeed, altering L2 learners' attitudes towards CF has been proven possible (Plonsky & Mills, 2006). The fact that attitudes towards CF are responsive to treatment

encourages further research into L2 learners' perceptions and preferences because attitudes that are not conducive to learning can be changed.

Due to the significant effects that perceptions and preferences may have on the effectiveness of CF, a number of exploratory studies examined L2 learners' attitudes towards CF (e.g., Diab, 2005; Elwood & Bode, 2014; Hyland, 2003; Lee, 2005; Leki, 1991; Montgomery & Baker, 2007; Schulz, 1996; Wang, 2015). The common conclusions from these studies indicate an overall positive attitude towards CF and a preference for more frequent, detailed, and explicit CF. Most L2 learners use form-focused CF in subsequent draft revisions (Hyland, 2003), value repeated CF (Hyland, 2003), expect to receive CF on all writing errors (Lee, 2005; Leki, 1991), and prefer explicit and detailed written CF (Elwood & Bode, 2014).

Studies (Amrhein & Nassaji, 2010; Baker & Bricker, 2010; Lee, 2005, 2008; Leki, 1991) that took a closer look at L2 learners' perceptions and preferences of different CF strategies revealed some interesting learner concerns. These concerns, which appear to govern learners' perceptions and preferences of different CF strategies, include concerns about the time required to make use of CF, clarity and adequacy of CF for making correct revisions, promotion of learner autonomy, and CF conductivity to learning. These concerns were revealed through interviews with learners following the completion of surveys about their perceptions and preferences of CF. Some of these concerns are actually at odds with each other, and their interplay may explain differences between learners' perceptions and preferences. For instance, learners who have a positive perception of clues and directions as a form of CF stated that clues and directions promote learner autonomy and are more likely to lead to remembering and learning while learners who had less positive perceptions stated that this form of CF is not adequate and

sufficient to lead to correct revisions (Amrhein & Nassaji, 2010). A similar interaction between different concerns can be seen in L2 learners' perceptions and preferences of CF.

Two CF studies (Amrhein & Nassaji, 2010; Leki, 1991) surveyed L2 learners about their perceptions of the usefulness of different written CF strategies provided by teachers. Four CF strategies were common between these two studies: refer to a specific section in a grammar handbook, label error type, correct error directly, and identify error visually. As presented in Table 7, results of both studies are somewhat at odds with each other. Specifically, Leki identified positive perception of handbook referencing while Amrhein and Nassaji identified neutral perception. In contrast, direct correction was positively perceived in Amrhein and Nassaji and neutrally perceived in Leki. In fact, direct correction was the most positively perceived CF strategy in Amrhein and Nassaji while handbook referencing was the most positively perceived in Leki.

Table 7

Usefulness Perception of Six CF Strategies

CF Type	Leki, 1991 (5-point Likert scale)	Amrhein & Nassaji, 2010 (5-point Likert scale)
Refer to grammar handbook	$\mu = 4.22$, $sd = 1.05$	$\mu = 2.8$, $sd = NA$
Label error type	$\mu = 3.5$, $sd = 1.08$	$\mu = 2.6$, $sd = NA$
Arrow clue	$\mu = 3.23$, $sd = 1.35$	
Direct correction	$\mu = 3.39$, $sd = 1.34$	$\mu = 4.1$, $sd = NA$
Identification	$\mu = 2.22$, $sd = 1.19$	$\mu = 2.4$, $sd = NA$
Direct correction + label error type		$\mu = 4.1$, $sd = NA$

There are no key differences between learner populations in the two studies that would explain this difference in perceptions except for educational settings. Leki sampled participants from freshman English in non-native speakers' classes while Amrhein and Nassaji sampled from private English-language schools with communicative-orientation. This difference in educational settings may have led to differences in learners' perceptions between the two studies. It is possible that freshman L2 learners favorably perceive handbook referencing because they are at an advanced proficiency level and their grammar errors tend to be few. This would give them the confidence to consult their handbooks and successfully self-correct their own errors. Further, few grammars errors mean less time spent on what can be considered a long error correction process. On the other hand, lower proficiency learners in communicative-oriented classrooms could favorably perceive direct correction because it does not distract them too much from focusing on fluency and expressing meaning.

Differences in usefulness perceptions discussed above indicate that perceptions may be mediated by diverse learner concerns. These concerns could include concern about the time required to understand and apply CF correctly, concern about interruption of ideas, and concern about the manageability of large and frequent CF. Identifying these concerns and exploring how they interact with each other to shape learners' preferences and perceptions is important for providing CF that is pertinent to learners' needs. However, these concerns have not been thoroughly explored and more research is required to reach conclusive findings.

The research findings reviewed so far pertain to CF from teachers in classroom settings. It remains to be seen if these findings generalize to learners attitudes towards automated CF in AWE settings as well. Up until now, studies (e.g., Wang, 2015) that examined CF in AWE tools focused on learners' perceptions of the particular CF strategy employed by those tools and

therefore did not reveal how learners would perceive other alternative CF strategies in comparison. The present study is mainly motivated by the lack of research that directly examines learners' perceptions and preferences of CF in AWE settings in addition to examining learners' perceptions and attitudes towards the novel approach of automated graduated CF.

2.7. Research Questions

The literature review above has identified the important role of CF in promoting L2 morphosyntactic development and L2 accuracy. It explored how the effect of CF is mediated by different CF strategies in both oral and written L2 production. It underscored the need for further research into the long-term impact of different CF strategies and mediating factors. It identified methodological limitations that undermine the findings of many CF studies. It summarized the findings of CALL-based CF studies. Further, the previous section reviewed studies that explored the use of graduated CF as a novel CF strategy in classroom settings. Finally, it identified the lack of research about L2 learners' perceptions and preferences of CF, different CF qualities and strategies, and the use of graduated CF in AWE settings.

The present study addresses this literature gap by developing an online writing tool that provides automated graduated CF. It uses this tool to introduce graduated CF to L2 learners and explores their perception of graduated CF through stimulated recall interviews. Further, it employs multiple data collection methods, including surveys and interviews, to explore how L2 learners perceive other forms of CF in comparison to graduated CF.

The following questions are addressed by the present study:

- How do intermediate to advanced L2 learners perceive automated written CF and what are their preferred CF frequency and timing?

- How do intermediate to advanced L2 learners perceive different automated written CF strategies in regards to their desirability, clarity, and usefulness?
- What effects do usage frequency and past experience with AWE tools have on intermediate to advanced L2 learners' preferences and perceptions of automated CF and its different strategies?

CHAPTER 3. METHODOLOGY

This chapter presents the research methodology employed to investigate the questions posed in this dissertation. Being descriptive, the present study employed surveys and interviews for collecting quantitative and qualitative data about learners' preferences and perceptions of CF and its different qualities and strategies. The following sections present and discuss the design, target structure, participants, materials, as well as the data collection and analysis procedures of the present study.

3.1. Design

This study employed a mixed-methods design to explore L2 learners' perception, preferences, and past experience with different automated CF strategies in comparison with automated graduated CF. Johnson and Onwuegbuzie (2004) define mixed-methods research as "the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study" (p. 17). Thus, mixed-methods research is capable of capturing a bigger picture of the phenomenon under investigation because it does not restrict researchers to a limited set of research methods, techniques, and explanatory approaches. In mixed-methods research, researchers have the choice to select from both qualitative and quantitative methods and techniques that best suit their needs and contribute additional insights into their investigations.

For this reason, the present study employed both surveys and interviews to explore L2 writers' perception, preferences, and past experiences with different automated CF strategies. Specifically, survey data is used to determine common factors affecting L2 writers' perceptions and preferences. Moreover, the data is used to determine if their perceptions and preferences are mediated by past automated CF experience. On the other hand, interview data is used to

determine the underlying reasons behind the perceptions and preferences of automated CF strategies that L2 writers expressed on the surveys. These two data types should lead to deeper understanding of L2 writers' perceptions and preferences and help to reach effective real world implications. As such, knowledge of how L2 writers perceive and what they prefer about CF should help optimize the presentation and delivery of automated CF strategies.

The expansion and initiation models of mixed-method design were employed in the present study (Greene et al., 1989). First, expansion helps in extending the breadth and range of investigation by including multiple data types through different research methods. In the present study, expansion is evident in the inclusion of survey and interview data in order to extend the coverage of investigation. Thus, this study answers both what and why questions about perceptions and preferences of different CF strategies. Second, initiation helps in reaching new perspectives and understandings of the phenomenon under investigation by "the recasting of questions or results from one method with questions or results from the other method" (Greene et al., 1989, p. 259). In the present study, initiation helps in reaching new perspectives and understandings of perceptions and preferences of different automated CF strategies by conflating survey data with interview data.

3.2. Target Structure

A single grammatical structure was chosen for the present study in order to have participants go through the same feedback experience and minimize interference from variation in target structure type: the simple past tense. This structure was the target of investigation in a number of past CF studies due to exhibiting delayed acquisition by L2 learners, having a regular-irregular dichotomy, and being responsive to CF treatment (Dulay & Burt, 1974; Ellis et al., 2006; Kartchava & Ammar, 2014b; Makino, 1980; Monteiro, 2014; Yang & Lyster, 2010).

Although these characteristics of the simple past tense may not be relevant to the descriptive nature of the present study, targeting the simple past tense could offer additional insights to the findings of past experimental CF studies targeting the same structure by bringing a different perspective.

3.3. Participants

Initial attempts to recruit participants at a low proficiency level did not yield enough participants to conduct statistical analysis of survey data. Therefore, ESL speakers at any proficiency level were recruited for the study. A total of 30 ESL speakers from different cultural, national, and L1 backgrounds were recruited. Many were taking ESL classes while others were taking college-level classes. It was assumed based on the educational levels and the intensive English program's levels of participants that they were at intermediate to advanced proficiency levels. All participants were living in a small college town in the United States at the time of data collection. In order to encourage participation, a monetary incentive was offered to all participants. Specifically, each participant was offered 5 US dollars and was entered into a drawing for 200 US dollars. The diversity of L1 backgrounds and proficiency levels of participants means that the findings would not capture variations in preferences and perceptions mediated by differences in these variables. In other words, the findings would only capture general preferences and perceptions that are subject to unexplored variation in L1 background and L2 proficiency level.

3.4. Materials

Two types of materials were used in the study. The first type was for introducing and familiarizing participants with automated graduated CF. This was accomplished by developing an AWE tool that provided graduated CF on written errors. The second type of material was for

collecting data about participants' preferences and perceptions of different CF strategies, including an automated graduated CF strategy as experienced in the AWE tool. It consisted of surveys and semi-structured and video-stimulated interviews.

3.4.1. AWE Tool

This study explored L2 writers' perceptions and preferences of different automated CF strategies in comparison with automated graduated CF strategy. However, automated graduated CF was not offered by any commercial or free AWE tool available on the market today. This had motivated the author to design and develop an online AWE tool called *wTutor* (standing for *Writing Tutor*) that delivers graduated CF. Figure 7 shows a screenshot of *wTutor*.

wTutor provides participants with graduated CF in an AWE environment. Delivery of graduated CF proceeds as follows. First, *wTutor* presents participants with a short descriptive text inside a text editor. This text contains many instances of ungrammatical past simple use. Most participants if not all are very familiar with past simple tense. Therefore, participants were asked to use the tool as if they are not familiar with simple past tense. A question placed at the top of the editor gives context to the text and makes evident the timeframe mismatch between the events and how they are described. Second, participants click on an icon of a magnifying glass located at the top left corner of the editor. This initiates automated error identification, analysis and delivery of graduated CF.

wTutor is designed to identify English regular and irregular simple past tense errors in learner writings and provides graduated CF. It consists of five modules: text editor, analysis module, feedback module, logging module, and survey module. These modules are described below.

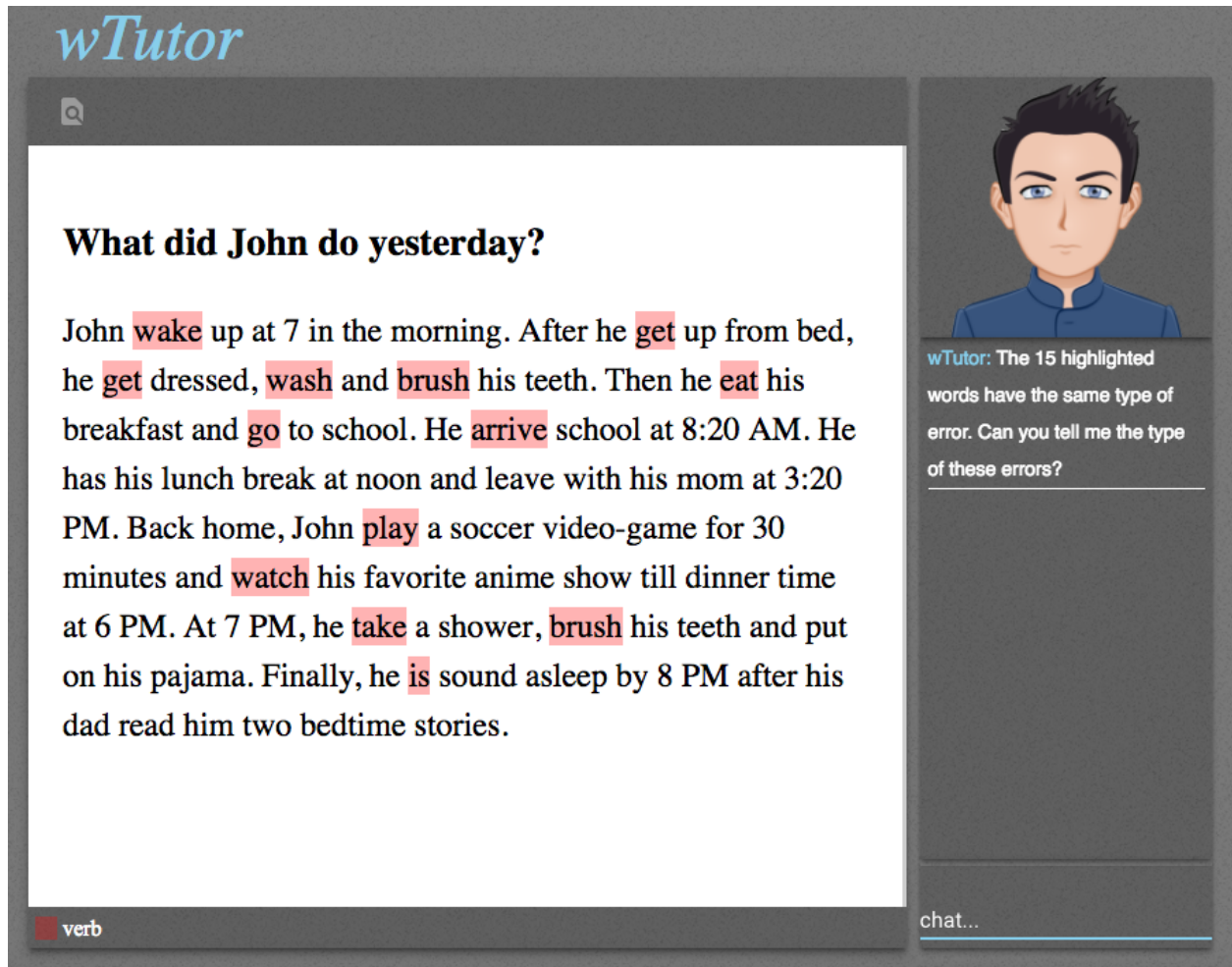


Figure 7. A screenshot of wTutor.

3.4.1.1. Text editor.

The text editor consists of an area for entering text and a button for requesting automated text analysis. For the purpose of the present study, entering text into the editor is disabled and the text editor is pre-populated with a prompt and paragraph answering that prompt. The prompt asks, “What did John do yesterday?” This way it sets the appropriate timeframe and construction that should have been used for the pre-populated written response.

3.4.1.2. Analysis module.

The analysis module uses CyWrite Analyzer for parsing and detecting errors in submitted texts. CyWrite Analyzer is a natural language processing engine, developed by Evgeny

Chukharev-Hudilainen (Chukharev-Hudilainen & Saricaoglu, 2014; Feng et al., 2016), which uses “the statistical parser from the Stanford CoreNLP suite to analyze the syntactic structure of sentences, and then applies formal rules to identify and extract various linguistic features of interest” (Chukharev-Hudilainen, n.d.). For the purpose of the present study, this module classifies all grammatical constructions other than the simple past tense as errors. This is because the appropriate timeframe and construction that should have been used in the text pre-populating the editor are set by the prompt above the text.

3.4.1.3. Feedback module.

The feedback module is responsible for receiving errors detected by the analysis module and preparing and delivering feedback. The tool offers two types of CF strategies: graduated tutor-based CF and non-graduated metalinguistic CF. However, the present study only employs graduated CF strategy. Therefore, it is described next in more detail.

Table 8

Description of Dialog States

State	Description
Reporting Error	This is the first tutor state in the graduated CF dialog. In this state, the computer tutor highlights all instances of one common error type and communicates to participants that the highlighted words share a common error type. Then, it asks participants if they know the error type shared between highlighted words. If participants responded with negation, the tutor transitions to Explaining Error Type state. On the other hand, if participants responded with affirmation, the tutor transitions to Evaluating Knowledge about Error Type state. (see Appendix I. 2. for a screenshot of reporting error state)
Explaining Error Type	In this state, the computer tutor explains the error type through a series of questions in order to provide the least amount of help to participants to understand the error type. The tutor provides hints when incorrect responses are entered and supplies correct answers if participants type incorrect answers after hints are given. For example, for the simple past tense error type that is the target structure in the present study, the tutor

Table 8
continued

	<p>first points with an arrow at the first sentence with simple past tense error and tells participants that they have made a simple past tense error. Then, it points at the main verb in the sentence and tells participants that that is the main verb in the sentence. Afterward, the tutor asks participants about the current tense of the main verb. If an incorrect answer is entered, the tutor asks participants if the current verb is in the present or past tense. Then, the tutor asks participants about the timeframe of the activities described in the pre-populated text. If participants enter the correct timeframe (i.e., past), the tutor asks them to change the main verb to simple past tense. (see Appendix I. 3. For a screenshot of explaining error type state)</p>
Evaluating Knowledge about Error Type	<p>The computer tutor enters this state when participants indicate that they know the error type of highlighted text. The goal of this state is to verify if indeed participants know the target error type. To achieve this, the tutor asks participants for the name of the error type. If participants enter an incorrect error type or a name not recognized by the tutor, the tutor displays a list of error types for participants to select the target error type from. Then, the tutor displays a sentence with the target error type from the pre-populated text, underlines the main verb, and prompts participants to correct the sentence. If participants enter an incorrect verb form, the tutor gives them a second chance before supplying the correct verb. (see Appendix I. 4. for two screenshots of evaluating knowledge about error type state)</p>
Ending Dialog	<p>This is the last state in the dialog and comes after Explain Error Type or Evaluate Knowledge about Error Type states. In this state, the tutor prompts participants to continue correcting similar errors in the text. Then, the chat-box is dismissed. (see Appendix I. 5. for two screenshots of ending dialog state and the state of editor after dialog)</p>

Graduated CF is dialogic, graded, and contingent on learners' need. These specifications of graduated CF are accomplished in *wTutor* through automated dialogs with participants. These dialogs are initiated in response to simple past tense errors in the text pre-populating the editor. First, all simple past tense errors are color highlighted. Then, a chat-box appears next to the text editor with a cartoonish avatar, script box, and an input box for participants to type their responses. The computer tutor communicates with participants through sound and text while

participants communicate with the tutor through text only. The computer tutor interacts with participants through a series of questions in order to lead them to correct highlighted grammar errors with the least amount of help. Appendix I. shows screenshots of the tool delivering CF through dialog.

The computer tutor's graduated CF dialog is organized into four main states: reporting error, explaining error type, evaluating knowledge about error type, ending dialog. The computer tutor uses a state machine to transition between different states in the dialog. Table 8 provides a description of these states.

3.4.1.4. Logging module and survey module.

The logging module logs all interactions between participants and tutor, including participants and tutor dialog. On the other hand, the survey module is responsible for collecting survey data from users. It presents the survey just before participants exit the program.

3.4.2. Video-stimulated recall.

Video-stimulated recalls collect participants' reflection about their experience with automated graduated CF offered in *wTutor*. This data collection method should provide information about how participants perceive automated graduated CF and the different steps involved in the interaction between participants and the computer tutor.

Video-stimulated recalls use screen-recordings of participants' interaction with the tool to probe for their perception of *wTutor* and the steps involved in the interaction. These screen-recordings help focus participants' attention on the various steps in their interaction with the *wTutor* and collect information about how they perceive these steps. Four main questions guide video-stimulated recalls. These questions are asked during playback of relevant interactions on the screen-recordings. The four main questions are listed in Appendix A. However, recall

questions are not restricted to these four. Other questions may be asked during recall sessions for clarification purposes and to prompt participants to elaborate on some of their interesting ideas.

3.4.3. Survey.

The survey collects quantitative data about participants' perception, preferences, and experiences with automated CF. Specifically, it explores how participants perceive the usefulness of automated CF in general and what preferences they have concerning the timing and frequency of automated CF. Further, it explores how much desirable, useful, and clear participants see different automated CF strategies. The strategies investigated by the survey are identification, direct correction, metalinguistic CF, and graduated CF. Finally, the survey collects data about participants' past experience with AWE tools and rating of their experience with each tool.

The survey is part of *wTutor*. It was offered immediately after participants complete the video-stimulated recall activity in order to ensure that participants have fresh recollection of automated graduated CF and maximize the return rate of surveys. The survey shown in Appendix B. has 15 required items and one optional item. An online form of the survey was used in the actual study. Items 1, 2, and 3 ask students to indicate their level of agreement with statements about their perceptions of the usefulness of automated CF and their desire to receive it frequently. They responded on a Likert Scale from 5 (=Strongly agree) to 1 (=Strongly disagree), with a high score meaning strong agreement that automated feedback is useful. Item 4 enquires about the preferred timing of written automated CF using a five-point Likert Scale: Immediately (5), after done writing the sentence (4), after done writing the paragraph (3), after done writing the paper (2) and when I ask for correction (1). Therefore a high score indicates the desire for immediate feedback, and a low score means no desire for immediate feedback unless requested.

Items 5, 6, 7, and 8 enquire about the perceived desirability, clarity, and usefulness of the four automated CF strategies listed above using a Likert Scale from 5 (=Strongly agree / Very clear / Very useful) to 1 (=Strongly disagree / Very unclear / Very unuseful). Each one of these items included three sub-items for the three measured variables. Different CF strategies are visually illustrated to ensure that participants have a comparable understanding of these strategies. Items 9 and 10 enquire about past experience with AWE tools, namely, usage frequency and rating of used AWE tools. For these items, participants are instructed to consider word processors such as Microsoft Word and Libre/OpenOffice Writer as AWE tools because they provide automated CF. Further, Item 10 is optional because some participants may have never used an AWE tool in the past.

3.4.4. Semi-structured interview.

While the surveys collect information about participants' perceptions and their preferences of CF and its strategies, the interviews collect information about the reason that participants hold such perceptions and preferences. For this purpose, the surveys are used to prompt participants during one-to-one interviews to discuss their perceptions and preferences of CF in greater depth. In the interviews, participants are shown their completed surveys and asked about the selections they made. In addition to survey-guided prompts, participants were asked for their overall evaluation of the tool and if they see any weakness or advantage to providing CF through automated graduated dialog. Appendix C. lists the questions used in the interviews. All interviews are audio recorded for subsequent transcription and analysis.

3.5. Data Collection Procedure

Data was collected during individual meetings with participants. These meetings were held in two public libraries and the building of an ESL program. At the beginning of the meeting,

the author introduced participants to the research activities they need to complete. In the order of presentation, these activities are a) a 5-10 minutes test-trial of automated graduated CF via *wTutor*, b) video-stimulated recall of experience with *wTutor*, c) a survey, and d) a semi-structured interview. The total amount of time required to complete these activities was 30 to 45 minutes per participant. Prior to the beginning of each research activity, participants were introduced to that activity in greater detail.

For the first activity, participants were informed about the purpose of the tool and the way it provides CF. Then, they were instructed to use a laptop computer to interact with the tool. Their interaction with the tool was screen-recorded using Apple QuickTime Player. During the interaction, the author was standing by to assist with any technical issue and clarify any question participants may have about how to interact with the tool. As explained in section 3.4.1., participants' responses to the first tutor question lead them to one of two different graduated dialog branches. In order to have them experience both main branches of the graduated dialog, participants were asked to try automated graduated CF twice. In the first trial, they were allowed to respond freely to the chatbot in *wTutor*. But in the second trial, participants were instructed to give a specific response to the first question in order to experience the other graduated dialog branch.

Immediately after participants completed their trials of *wTutor*, their perception of automated graduated CF as presented in *wTutor* was collected using video-stimulated recall. First, they were introduced to the process of video-stimulated recall and were informed of the type of questions to expect. Then, screen-recordings of their interaction with the tool were played back and intermittently paused at key steps in the interaction. Participants were also allowed to

pause the playback where they want to share comments. These video-stimulated recall sessions were audio-recorded to facilitate later transcription and analysis.

In the third activity, participants were instructed to complete an online survey about their perceptions, preferences, and experiences with written automated CF. The survey required 10 to 15 minutes to complete. Participants were told that the first five items on the survey refer to automated CF in general because the author noticed that some participants thought these items refer to the specific CF strategy they have experienced in *wTutor*. Also, participants were informed that upon completing the survey they would be asked in an interview about their reasons for selecting specific choices on the survey. As participants were completing the survey, the author was available to respond to any question or clarification request that participants might have.

Semi-structured interviews were the fourth and last activity completed by participants. First, participants were reminded that their interviews would be recorded. Then, as explained in section 3.4.4., completed surveys were used to prompt participants for the specific reasons behind the kinds of perceptions and preferences they have shown towards automated CF. The interviews took 10 to 15 minutes to complete.

3.6. Data Analysis

The methodology outlined above produced transcripts of semi-structured interviews and video-stimulated recalls and a dataset of raw survey responses. This section describes the data analysis procedure and is organized according to the three research questions presented earlier.

3.6.1. Research Question 1.

The first research question explored ESL writers' perception of automated written CF and their preferred CF frequency and timing. Data addressing this question came from both surveys

and semi-structured interviews. Descriptive statistics were computed to identify the overall average perception of automated written CF and preferred CF frequency and timing. Then, an exploratory factor analysis was conducted to determine if patterns exist in the preferences and perceptions measured by the survey.

Results of statistical analysis of survey data were interpreted in light of the qualitative analysis of rich semi-structured interview transcripts. The transcripts were coded by the author using a grounded and open-ended coding procedure where codes were based on emergent and recurring patterns across multiple interviews. As outlined by Charmaz (2006), grounded theory approach to coding involves initial and focused coding phases. In the initial coding phase, each small segment of data will be labeled and categorized in relation to the analytic questions being asked. Then, the focused coding phase will use the most frequent codes to identify recurring themes and reach a credible understanding of participants' perceptions and preferences.

Results from both quantitative and qualitative analysis provided both a general picture of perception of automated written CF and preferred CF frequency and timing and more in-depth understanding of how and why variations exist if any.

3.6.2. Research Question 2.

Research question 2 explored how ESL writers perceive different automated CF strategies with special emphasis on automated graduated CF. Data addressing this question came from surveys, semi-structured interviews, and stimulated recall interviews. While descriptive statistics were reported for survey data to show the average of perception and preference responses, an exploratory factor analysis was conducted to determine if any latent relationships exist between perceptions and preferences of CF strategies. Then, open-ended coding and analysis of semi-structured interviews and video-stimulated recall interviews helped in

understanding the reason for the perceptions and preferences displayed by participants on the surveys.

3.6.3. Research Question 3.

Research question 3 explored if past experience with automated CF influences L2 writers' perceptions and preferences of automated CF and different automated CF strategies. Analysis of survey data addressed this question. Similar to the analysis procedure of research questions 1 and 2, survey data was reported using descriptive statistics that shows the mean, median, and standard deviation of past CF experience responses. Then, exploratory factor analysis determined if a pattern exist between past CF experience responses and other survey items.

Exploratory factor analysis was performed using R and relevant R packages. Use of R in performing statistical analyses in the present study was motivated by Mizumoto and Plonsky's (2015) call for greater use of R in Applied Linguistics research. Reproducibility is one of the most important reasons among several put forward by Mizumoto and Plonsky for using R. Reproducibility here refers to the ability to recreate and replicate statistical analyses used in a given study. According to Mizumoto and Plonsky, R significantly facilitates reproducing statistical analyses because it is a code-based terminal-driven program. This means that when datasets are shared with the wider research community for the purpose of reproducing primary studies, R codes used for reaching their findings can be shared as well.

3.7. Chapter Summary

This chapter described the research methods employed to assemble a detailed picture of L2 writers' perceptions and preferences of automated CF. As explained in this chapter, a mixed method design was opted for in order to reach an in-depth understanding of L2 writers'

perceptions and preferences of automated CF. For this purpose, the present study collected three different data types: surveys, semi-structured interviews, and video-stimulated recall interviews. Analysis of these data types involved calculating descriptive statistics, performing exploratory factor analysis of surveys, and coding the semi-structured and video-stimulated interviews using grounded theory approach. The next chapter presents the results of data collection and analysis procedure outlined in the chapter.

CHAPTER 4. RESULTS AND DISCUSSION

The research methods outlined in the previous chapter gather data addressing the three questions advanced by the present study. These questions explore perception of automated written CF and its different strategies, CF timing and frequency preferences and the interplay among CF perceptions and preferences and past automated CF experience. The data collected to answer these questions are surveys, semi-structured interviews, and video-stimulated recall interviews. The present study collected a total of 30 surveys, video-stimulated interviews, and semi-structured interviews.

Prior to conducting statistical analysis, survey data was reviewed for missing responses and checked for having acceptable internal item consistency coefficient value and meeting the factorability assumptions of exploratory factor analysis. First, initial review of survey responses led to excluding item 10, which enquires about AWE tools used by participants due to missing responses. This item was intentionally not required because it is very likely that some participants may not have used AWE tools in the past. Second, Cronbach's alpha was computed using R package 'psych' version 1.6.4 to estimate internal item consistency. The coefficient alpha is at an acceptable 0.74 value indicating that between-subject variability is greater than within-subject variability. Third, two factorability assumptions of exploratory factor analysis were assessed through Kaiser-Meyer-Olkin's (KMO) Measure of Sampling Adequacy (MSA) and Bartlett's Test of Sphericity. KMO returned an overall MSA of 0.44, which is below the minimum acceptable value of 0.5 (Cerny & Kaiser, 1977; Kaiser, 1974). For this reason, MSA was recalculated several times with one unique survey item removed each time to determine if removing one survey item raises MSA value above the minimum threshold. This step determined that only removing item 8a raises MSA to 0.51. Item 8a enquires about the desirability of

graduated CF by L2 learners. The KMO value of .51 and the statistically significant Bartlett's Test of Sphericity ($\chi^2 = 238.2672$, $p < 7.856561e-10$) confirm the suitability of data for exploratory factor analysis.

The next three sections present and discuss the results of surveys, semi-structured interviews, and video-stimulated recall interviews as they pertain to the three research questions of the present study. The results of factor analysis are presented first because they pertain to all three research questions.

4.1. Factor Analysis Results

After completing the pre-processing of survey data, a scree test was conducted on the data to determine the number of factors to extract. This test determined that 3 meaningful factors can be extracted from the surveys. Figure 8 shows a screeplot of eigenvalues of factor analysis as outputted by R. Eigenvalues above the red line are optimal factors to extract.

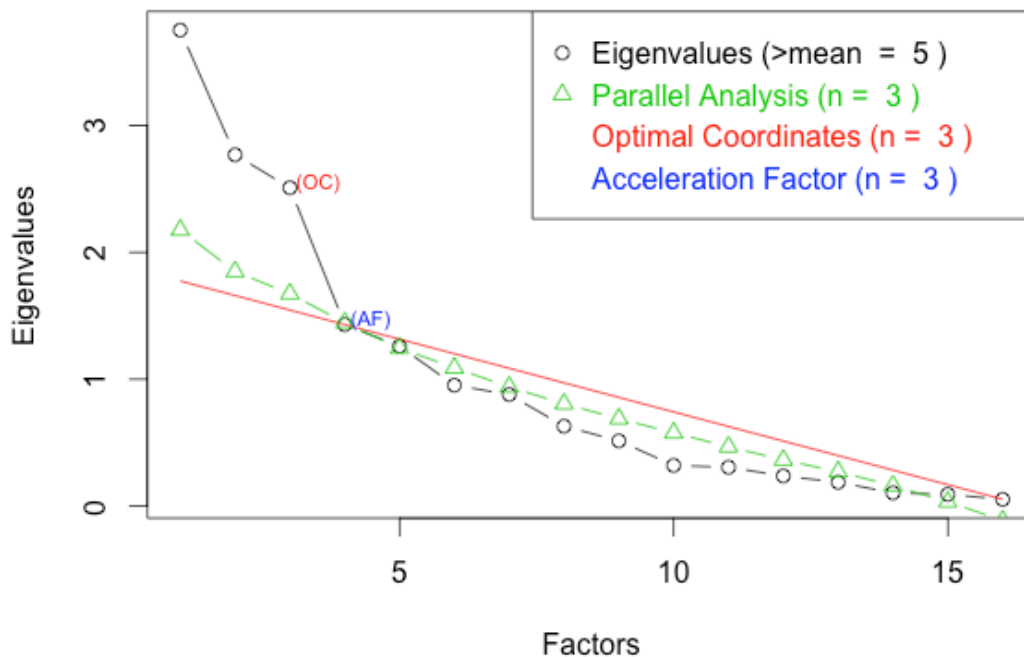


Figure 8. Screeplot showing the number of optimal factors to extract from surveys.

Factor analysis of surveys produced three factors accounting for 0.48 of total variance as shown in Table 9. The total variance value shows that 48% of the variation in survey responses is accounted for by the three identified factors. All three factors have a relatively high internal reliability as measured by Cronbach's Alpha, .80, .81, and .71, respectively. Kline's (1994) cut-off points for factor loadings are used for classifying loading values into high (above 0.60) or moderately high (above 0.30) loadings. Factor 1 explains 18% of total variance and mostly consists of items about the clarity and usefulness of explicit CF (Items 6a, 7, 7a, 7b, 8b). Factor 1 also includes item 4, which is concerned with CF timing but with a lower loading. Based on the high positive loadings of clarity, desirability, and usefulness items, factor 1 is labeled *Desire for Clear and Useful CF*. Factor 2 explains 16% of total variation and constitutes of items 2, 3, 6, and 6b. These four survey items measure preferred frequency of CF and perceptions of desirability and clarity of direction corrections. Thus, Factor 2 is labeled *Direct correction with Comprehensive CF*. Finally, Factor 3 explains 14% of total variation and consists of items 5.a., 5.b., 5.c., and 9. The first three items measure desirability and perceptions of clarity and usefulness of identification alone as a CF strategy. The fourth item measures how frequently participants use AWE tools. Thus, Factor 3 is labeled *Identification and Frequency of AWE Use*. Appendix J. shows a plot for exploratory factor analysis results, which is useful for understanding how survey items load with other potential factors.

Table 9

Results of Exploratory Factor Analysis

Item	Factor 1	Factor 2	Factor 3
<i>I. Desire for Clear and Useful CF</i>			
4. When I make an error, I want to be corrected.. (timing)	.37		

Table 9 continued

6. a. How clear is this type of feedback? (direct)	.51	.31
7. I want the computer to highlight my grammar errors and tell me their types. (meta CF)	.85	
7. a. How clear is this type of feedback? (meta CF)	.70	
7. b. How useful is this type of feedback? (meta CF)	.84	
8. b. How useful is this type of feedback? (graduated CF)	.59	.54
II. Direct Correction with Comprehensive CF		
2. I want to be corrected on all types of grammar errors.	.56	
3. I want to be told about every grammar error in my writing.	.63	
6. I want the computer to highlight my grammar errors and tell me the correct forms. (direct correct)	.71	
6. b. How useful is this type of feedback? (direct correct)	.92	.36
III. Identification and Frequency of AWE Use		
5. a. I want the computer to only highlight my grammar errors. (identify)		.80
5. b. How clear is this type of feedback? (identify)	.43	.73
5. c. How useful is this type of feedback? (identify)		.68
9. How often do you use automated writing evaluation software that show and correct your errors?		-.37

4.2. Research Question 1: CF Perception and Timing and Frequency Preferences

The first research question asked how ESL learners perceive automated CF and what CF timing and frequency they prefer. Survey items 1, 2, 3, and 4, and semi-structured interviews addressed this question. Responses to the four survey items are summarized in Table 10 and Table 11. To facilitate comparison due to the small number of participants ($n=30$), responses with *Strongly*

agree were combined with *agree* and responses with *Strongly disagree* were combined with *disagree*.

Table 10

Summary of Responses to CF Perception and Preference Statements on a 5-Point Likert Scale (n=30)

Item	Agree / Strongly agree	Neutral	Disagree / Strongly disagree	Mean	SD
1. CF usefulness	28 (94%)	1 (3%)	1 (3%)	4.17	0.65
2. CF for all error types	27 (90%)	0	3 (10%)	4.40	1.04
3. CF for all errors	28 (93%)	0	2 (7%)	4.53	0.82

Table 11

Summary of Responses to Timing and Frequency Preference Statements on a 5-Point Likert Scale (n=30)

Item	when I ask for correction	after done writing the paper	after done writing the paragraph	after done writing the sentence	Immediately	Mean	SD
4. CF timing	6 (20%)	2 (7%)	7 (23%)	8 (27%)	7 (23%)	3.27	1.44

Survey item 1, “*I learn from error corrections how to use grammar rules correctly,*” measured learners’ perceptions of the usefulness of automated form-focused written CF for developing their ability to correctly apply grammar knowledge in their writings. Most responses were positive with 94% of participants choosing to agree or strongly agree that automated form-

focused CF has a positive effect on correct language use. Of the remaining participants, 3% indicated a neutral opinion and 3% indicated disagreement with survey item 1.

In light of earlier CF studies (e.g., Diab, 2005; Elwood & Bode, 2014; Hyland, 2003; Schulz, 1996), this overwhelmingly positive attitude towards the value of CF for promoting L2 development is not surprising. These studies have identified a very positive perception of form-focused CF. Response to item 1 shows that similar positive perception exists for automated CF. However, Elwood and Bode (2014) pointed that perception may be mediated by learners' L2 proficiency level. They found that low proficiency level learners have a less positive perception of CF than high proficiency level learners. In the present study, the small number of participants ($n=30$) and convenience sampling method used make it difficult to determine if L2 proficiency level mediates perception of CF.

Similar positive perception of automated form-focused CF was identified by Wang (2015) in her study of L2 learners' perception of automated CF in *Criterion*®. In contrast, Chen and Cheng (2008) found that perception of automated CF may be mediated by how and when automated CF is used. Specifically, they found that automated CF is perceived more positively if it was provided during early drafting and revision processes and with conjunction with teacher and peer CF. The current survey was not designed to assess the effect of the writing process and teacher and peer CF on the perception of automated CF. However, these are important factors that should be considered to promote the positive perception of automated CF. Interview data shed light on other mediating concerns of equal importance as will be reported next.

Semi-structured interview data were examined in order to explore participants' motivations for indicating this general positive attitude towards automated form-focused CF. Interview data identified a wide range of concerns and motivations behind participants' self-

reported perceptions. Some participants ($n=10$) shared that they consider automated form-focused CF useful because it brings their attention to errors and error types. In their opinion, this awareness should help in learning how to use grammar rules correctly. The following two quotes provide an example of this perception of automated CF:

I think it shows you where you are good at and where you are lacking, and when you see that you... maybe you notice something that you are missing in your grammar and then it's that much... in that regard kinda effective to note... to kinda push you (Participant 1, semi-structure interview)

Because they tell you are wrong and sometimes I cannot find the error so its that's good to let you know I am still wrong here (Participant 2, semi-structure interview)

In line with this notion that awareness helps in learning, some participants indicated that CF reminds them of correct grammar usage ($n=3$), helps in avoiding similar future errors ($n=5$), improves writing quality ($n=1$), pushes them to think about their errors and how to correct them ($n=1$), and boosts confidence in the quality of their own writing ($n=1$). These perceptions are illustrated by the following quotes:

sometimes I forget some of the grammar rules, so when the tool or software, or sometimes I need someone to remind me, for example some basic grammar, sometimes not only the basics sometimes exclamation points something like that. (Participant 3, semi-structured interview)

Actually I have been through this myself when I ah... When I like do one mistake, grammar or spelling mistake then the software correct that for me that then I will not do this, the same mistake again. So, it's used for to use the grammar corrections as like learning tool. (Participant 4, semi-structured interview)

I be confident that I didn't make any mistake (Participant 5, semi-structured interview)

A few past studies (Amrhein & Nassaji, 2010; Lee, 2005) have identified similar explanations for the positive perception of CF by L2 learners. Amrhein and Nassaji (2010) found that L2 learners believe form-focused CF reminds them of correct grammar usage, causes them

to reflect on their errors, and teaches them how to fix them. Further, Lee (2005) found that positive perception of CF by L2 learners is connected with L2 learners' beliefs that CF helps them know their errors and avoid similar errors in subsequent writing. The present study corroborates the findings of Amrhein and Nassaji (2010) and Lee (2005) by identifying similar beliefs towards automated CF as reported above.

Besides these beliefs about the value of CF for learning, participants shared a range of concerns about automated form-focused CF. These concerns relate to the notion of learner autonomy, time, and the accuracy of automated error detection. First, two participants indicated that automated CF should promote self-reliance of L2 learners as it helps them rely less on L2 instructors for error identification and correction as illustrated by this example quote:

I don't have to ask somebody whose native language is English for my essay and correct the grammar errors so it would be nice if I can do it myself with help of the software.
(Participant 6, semi-structured interview)

But, in contrast, one participant shared a concern about L2 learners' overreliance on automated CF for error detection and correction and not learning by attempting to identify and correct errors by themselves as quoted here:

Participant 7: Particularly, sometimes I don't want to correct and they correct me ... They don't even like give you options
Interviewer: Okay, I see your point. And you don't want that part. You want to take care of it on your own
Participant 7: Because I don't learn (Participant 7, semi-structured interview)

Another participant showed a concern about the time needed to receive and apply CF and doubted the value of CF when L2 learners are pressed for time to complete their writing assignments. Finally, one participant showed concern about frequent inaccurate automated error detection and believed that this would diminish potential benefits from automated CF. These two concerns are illustrated by the following two quotes:

Depends like sometimes you are rushing so you don't really have that time to think about what happened and I mean those <not clear> it depends like I am using a computer time it just do it. (Participant 8, semi-structured interview)

Interviewer: Is automated error correction helpful?

Participant 9: No, sometimes they make mistakes ... In spelling they catch the mistakes but grammar, Aaah, subject and verb they don't much sometimes (Participant 9, semi-structured interview)

Concerns about learner autonomy and time revealed by survey item 1 and follow-up interviews confirmed findings of past CF studies (Amrhein & Nassaji, 2010; Lee, 2005, 2008). Amrhein and Nassaji (2010) found that some L2 learners are concerned about the effect of CF on learner autonomy. Their findings also indicate that concern about learner autonomy may be mediated by CF strategy type and CF frequency. Similarly, Lee (2005; 2008) found that some L2 learners are concerned by learner autonomy, but the majority prefer to rely on L2 teachers for error identification and correction. Despite beliefs about the importance of learner autonomy, Lee found that the majority of L2 learners indicated a preference for more direct and comprehensive CF from their teachers. The mediational effect of CF strategy type and CF frequency on learner autonomy concern is further discussed when reporting the results of corresponding survey items below.

Concern about the accuracy of automated error detection reported by participants has also been reported in a number of past studies (Li et al., 2015; Lim & Kahng, 2012; Wang, 2015). This is not surprising given that automated CF are prone to detection and classification errors and L2 learners are well-aware of such technical limitations. Furthermore, this concern is not unique to automated CF. Colpitts (2016) reported a similar concern about the accuracy of peer-to-peer CF. The accuracy of peer-to-peer CF depends on peer's knowledge about the relevant target language and their proficiency level. Although concern about the accuracy of automated

CF is not common, it is still important to determine how this concern may affect desirability and use of automated CF. However, this topic is beyond the scope of the present study.

Survey item 2 and 3 enquired about L2 learners' desire "to be corrected on all types of grammar errors" and "to be told about every grammar error in [their] writing." Item 2 sought to determine if L2 learners prefer to receive CF on specific types of grammar errors because different types of grammar errors may have different levels of severity and effect on meaning. In contrast, item 3 sought to determine if L2 learners prefer to receive CF on some errors of a given grammar error type in order to be less overwhelmed with CF and to depend more on themselves in identifying and correcting similar errors.

The majority of responses to item 2 (90%) and item 3 (93%) were in favor of more comprehensive CF. A similar preference for comprehensive CF was reported in non-automated CF settings with 80% to 90% of L2 learners preferring more comprehensive CF from their teachers (Amrhein & Nassaji, 2010; Jeon & Kang, 2005; Lee, 2005, 2008). L2 learners prefer to be corrected on all errors in both automated and non-automated CF settings. Few L2 learners desire selective CF that only focuses on major error types. This preference shows how important and useful CF is to L2 learners. Interview data are reviewed next to determine why L2 learners prefer comprehensive CF.

In the interviews, the main reason cited by participants ($n=6$) for preferring comprehensive CF was that it helps them identify and correct all of their errors as illustrated by the following quotes:

... because I'm, I'm... I write e-mails or I write articles...Professional articles, professional Momento, etcetra... I need to correct all my [errors] (Participant 10, semi-structured interview)

I prefer to me that mistake to be corrected because may be I can't catch the mistake where it is but this will help me to know my mistake. (Participant 11, semi-structured interview)

I just need to know all my mistakes, mistake I am doing (Participant 12, semi-structured interview)

This reflects the importance of having accurate and error-free text for L2 learners. A similar explanation was also reported by Amrhein and Nassaji (2010) and Lee (2005) for the preference of non-automated comprehensive CF. Amrhein and Nassaji (2010) reported that L2 learners preferred comprehensive CF because they believe it leads them to write more accurate text. Similarly, Lee (2005) found that L2 learners prefer comprehensive CF because they want to have their text free of errors and they believe knowledge about all errors helps them learn and avoid making similar errors in the future. The present study confirms that similar beliefs about comprehensive CF exist in AWE settings.

In addition to this favorable perception of comprehensive CF, some participants voiced their concern about interruption of ideas caused by too much form-focused CF ($n=4$), concern about the accuracy of automated CF ($n=2$), and concern about overreliance on comprehensive CF by L2 learners resulting in less learning ($n=1$). These three concerns are illustrated by the following quotes:

If I receive all focus on grammar... it can interrupt my idea (Participant 13, semi-structured interview)

Participant 14: ...usually I wanted them to correct me for all all types of errors but sometimes I think because of the problems themselves I'm not smart enough, so sometimes they will correct me in some things that I don't think it's correct.

Interviewer: Okay, okay. So, you don't have confidence... enough confidence in the computer because it has some errors in correcting

Participant 14: Yes. Exactly. Detecting errors. (Participant 14, semi-structured interview)

... because sometimes if I don't know the simple past, I will check my grammar book, I will learn by myself, if this is so easy then I forgot it so easy, so I think either way, but I say Ok because it depends. Some students want immediate feedback, but some student intelligent student they would try to find another book. So I say I agree. It depends.
(Participant 15, semi-structured interview)

I think I don't want for all types because I feel some some some types are more important than the others. ...like for example, if I have umm... If I say maybe a run-on sentence error, I think for... yeah those kind of errors I think would be a bit serious than the others... than for example umm from something like a spelling error. And then for for spelling you know at least there are there are ways like Microsoft Word that could that could kinda correct my spelling errors, but for run-on or for more maybe which, which I call maybe a little bit serious, serious or global errors I think I want to be corrected but not for simple or or for errors that I kinda that could be corrected by other means.
(Participant 1, semi-structured interview)

First, a similar concern about comprehensive CF interrupting learners' focus on expressing their ideas has been reported in the literature. Lee (2005) reported that some L2 learners preferred selective CF because they believed selective CF is less overwhelming and therefore helps them maintain their focus on expressing their ideas and facilitates learning from errors. Likewise, a participant in the present study suggested that CF should be prioritized based on the severity of error types and their impact on meaning. The present study shows that a comparable amount of concern about comprehensive CF interrupting ideas exists in automated CF settings just like in non-automated CF settings.

Second, two participants indicated that their concern about the accuracy of automated CF makes comprehensive CF less desirable. They believe that due to technical limitations that make automated CF prone to errors, CF should be more selective. This is the second time concern was voiced about automated CF. The first time was during interviews about the desirability of CF. The reemergence of concern about the accuracy of automated CF underscores the significance of

this concern for some L2 learners. Past CF studies did not find this particular concern in connection with comprehensive CF, whether automated or non-automated.

Third, a participant voiced concern about overreliance on comprehensive CF for error detection and correction. For this participant, selective CF that focuses on major errors and key error instances pushes L2 learners to be more active in identifying and correcting their own errors and learn from this process. Lee (2005) has reported similar beliefs by L2 learners who prefer selective CF, namely that selective CF is better for learning. However, this is a rare view among participants interviewed in the present study and interviewed in Lee's study.

The three concerns reported and discussed here are recurring throughout the interviews. This underscores the importance of identifying, understanding, and addressing these concerns. Next, results of factor analysis that pertain to CF frequency items (i.e., item 2 and 3) are examined.

Based on Table 9, preferences for both comprehensive CF and direct correction have high positive loadings in Factor 2, *Direct Correction with Comprehensive CF*, with a factor loading of 0.56 and 0.63 for survey item 2 and 3 respectively. This shows a strong preference for direct correction with comprehensive CF. This preference can be explained in light of some of the concerns voiced by participants about automated CF, namely concern about the time required to process and apply CF and concern about CF interrupting focus on meaning. Direct correction requires the least amount of time for correcting errors in comparison with other CF strategies by directly supplying learners with correct forms. This quality of direct correction could make comprehensive CF more manageable and less overwhelming. By requiring less time for processing and applying CF, direct correction interrupts learners' focus on meaning less than other CF strategies.

By mediating concerns about time and interruption of meaning expression, direct correction becomes more preferred with comprehensive CF, which in itself may require more time and may cause a greater distraction. Combining direct correction with comprehensive CF creates a balance between competing concerns. Participants are concerned about the quality of their writing and desire an error-free text. However, at the same time, they do not want to spend too much time correcting surface-level errors and being distracted away from expressing their ideas. Consequently, combining direct correction with comprehensive CF appears to strike a balance between these competing concerns.

The discussion above examined the link between direct correction and comprehensive CF that was detected by factor analysis of the surveys. No such link was examined or identified in past studies and therefore generalizing this finding to other learner populations is not possible. Future studies need to directly examine the link between direct correction and comprehensive CF via quantitative means in order to determine the extent and nature of this link. Next, the result of survey item 4 about preferred timing of CF is reported and discussed.

Survey item 4, *“When I make an error, I want to be corrected.. (a) when I ask for correction, (b) after done writing the paper, (c) after done writing the paragraph, (d) after done writing the sentence, (e) immediately,”* enquired about learners’ preferences between five timings for receiving CF. Results show that most participants prefer to receive CF close to the time when the error was made: immediately (23%), per sentence (27%), and per paragraph (23%). Only 7% of responses were in favor of receiving CF after completion of writing task while the remainder 20% of responses favored receiving CF on-demand.

Attempts to draw on the literature for comparison with the above results about preferred written CF timing did not succeed. This is due to the nature of non-automated written CF,

namely that it generally does not allow for immediate feedback (Ellis, 2009b). For non-automated CF, learners' writings have to be collected, marked, and returned by teachers before learners get to receive CF. Although it is possible for teachers to deliver non-automated CF immediately through electronic means, it is not feasible due to time and logistical constraints. On the other hand, the timing of spoken CF has been examined in past studies because it easily allows for immediate and delayed timings. Interestingly, it was found that L2 learners prefer to receive spoken CF immediately after they finish speaking, which is similar to the preference for immediate written CF in the present study (Park, 2010). This indicates that similar beliefs and attitudes towards CF both spoken and written may be shaping learners' preferences of CF timing. Next, beliefs and attitudes about preferences of CF timing are examined and discussed.

Participants' explanation of their preferences of CF timing revealed a wide range of concerns and beliefs. Due to the short gap in time between error and correction, some participants believe immediate CF is easier to focus on and understand ($n=1$), helps in learning from errors ($n=2$), improves immediate subsequent writing by preventing similar errors ($n=2$), and keeps writers more focused on their writing ($n=1$). These positive perceptions are illustrated by the following quotes:

immediately it's easy to see and understand and get to the next sentence and go forward (Participant 5, semi-structured interview)

I'm one of those people when corrected immediately I remember better than when I write and write and at the end I find out my mistakes. I would not benefit a lot. (Participant 16, semi-structured interview)

if i get like for example the immediate correction that will help to avoid the mistake during my writing for complete paper so I think it is going to be helpful rather than you know wait till the end of writing the paper and see all the correction (Participant 17, semi-structured interview)

However, other participants were concerned about the interruption of ideas and distraction from meaning expression caused by untimely CF ($n=3$) while others preferred immediate CF only if it can be ignored and addressed later ($n=2$) as illustrated by the following quotes:

when I am writing and focussing on writing the ideas its better to just to write then correct what I have, Correcting the each and every single like word or sentence destroy my idea. (Participant 18, semi-structured interview)

I don't think it will affect [focus on meaning] because when I said, immediately it is highlighted, when I write it, if there is any mistake it just highlight it that's all. Then I have the option if I want to correct it immediately or I just leave it. (Participant 12, semi-structured interview)

Overall, participants appear to value immediate CF because it leads to identifying and addressing errors early, but they are concerned about it negatively impacting their focus on expressing ideas.

In contrast to immediate CF, receiving CF per sentence is preferred by some participants because it addresses their concerns about interruption of ideas ($n=3$), manageability of CF ($n=1$), writing accuracy ($n=2$), and accuracy of automated CF ($n=1$) as illustrated by the following quotes:

I don't like the correction in middle of writing. So I need to finish my idea first and then look back to the grammar. Sometimes while you are writing the computer come up and intervene your writing. So you have a problem here. It doesn't know. The computer doesn't know what I am wanting to write in the following. So maybe I will write one or two sentences and then look back for the previous sentences and make some edit and that one. Sometimes I need to edit maybe even the first sentence in the paragraph.... (Participant 3, semi-structured interview)

I think only when you finish the sentence maybe the computer will.. can figure out what kind of like.. a tense you should use if they told you immediately sometimes maybe maybe the word you don't want to use the computer want to tell you. (Participant 19, semi-structured interview)

[In] my experience, every time I write a sentence I'm confused whether I am right or not.
(Participant 20, semi-structured interview)

For three participants, receiving CF after sentence completion does not interrupt their focus on expressing ideas especially with those struggling with grammar and writing mechanics, because a sentence constitutes one complete idea. Further, one participant believes receiving CF per sentence does not overwhelm learners with CF compared with CF per paragraph or paper because a small and thus more manageable number of errors can occur in a sentence. Two participants also believe that receiving CF sentence-by-sentence will help improve immediate writing quality by preventing repetition of errors in subsequent sentences. Finally, one participant prefers receiving CF per sentence because a sentence is the minimum amount of text that can be reliably analyzed by computer for grammar errors.

Similar to CF per sentence, receiving CF per paragraph addresses L2 learners' concerns about interruption of ideas ($n=3$) and manageability of CF ($n=3$). Additionally, one participant shared that CF per paragraph addresses concern about time ($n=1$). The following quotes illustrate these perceptions.

Because the paragraph has the same information and if I fix sentence by sentence that's not really good writing.. editing I think. So, I should read paragraph and then correct the grammar and then correct the writing quality. [A paragraph represents] common information. (Participant 21, semi-structured interview)

because I don't choose the sentence because if I just write one sentence and give me a feedback, I think it's too, maybe I write the time is too long, maybe you can write the whole paragraph and you can tell me what mistake. I can check my mistake. But if I don't choose the write the paper, I think if you already write the whole paper you give me... maybe too much mistake for me. So I think a paragraph maybe five to six sentence I think is good for me to tell me. (Participant 22, semi-structured interview)

One thing is... I don't know about others, but for me when someone remind me about something I'll immediately put it in mind. So if I have something that make me... keep

thinking keep thinking I will stop there and I am stuck. Ok I am not making this mistake I am not making this mistake. So I will say after one paragraph is a 30 minute or 15 minute something like that and then they tell you ok <not clear> change the whole thing because the grammar error is not just one sentence sometimes it depends on the whole paragraph. So after I have better understanding Oh I did some error like that and the next paragraph I am not going to make that again instead of saying you're wrong think about that and you spend 5 minutes and another sentence like that. (Participant 8, semi-structured interview)

These participants believe that a paragraph expresses one complete idea and therefore receiving CF per paragraph does not interrupt their focus on meaning. They also consider the potential amount of CF given per paragraph manageable and non-overwhelming when compared with receiving CF at the end of a lengthy paper. Finally, one participant shared the belief that CF per paragraph saves more time spent in correcting errors compared with CF per sentence. For this participant, receiving CF per paragraph saves time by reading and applying CF for similar errors once per paragraph instead of sentence-by-sentence.

For the few participants ($n=2$) who preferred receiving CF after completing writing, interruption of focus on meaning and organization was the only concern behind this preference as illustrated by the following quote:

I think umm I want to move on and then be done with my writing because while writing it's not only grammar, it's at the same time the organization, the content. And then you're thinking at the same time while writing and then I think writing or grammar... you know is interfered by your thinking. And then whenever I get you know frequent umm... umm kinda correction I think it's kinda interrupting my thinking and then my... the flow of content in my writing. And then I think I feel first I should finish my writing, let's say it's, it's one page writing, one essay or maybe a couple of paragraphs and then I want to get corrected. (Participant 1, semi-structured interview)

These participants believe that form-focused CF should be delayed until ideas are fully expressed and text organization is finalized at the end of writing tasks in order to minimize interruption. Two more participants who indicated preference for receiving CF per paragraph

expressed similar beliefs about receiving CF at the end of writing. Similar concern over interruption of ideas was identified for participants who prefer receiving CF on demand. These participants preferred to trigger CF themselves because it allows them to prioritize other more important aspects of writing over form and accuracy.

Based on the results reported above, participants have four main concerns about different CF timings: time, manageability, focus on meaning, and writing accuracy. These concerns appear to shape their preferences between different CF timings. Participants prefer to spend less time on reading, understanding, and applying CF. They prefer to receive a manageable amount of CF. They prefer not to be distracted too much from focusing on expressing their ideas. They prefer to avoid repeating errors and to see more immediate improvement in writing accuracy. These concerns compete with each other. For instance, providing a manageable amount of CF means more interruptions. Based on survey and interview data about CF timing, it cannot be determined how the interaction between these opposing concerns shapes learners' preferences.

In line with past CF studies, the survey and interview responses analyzed so far show that most L2 learners favorably perceive CF as useful for L2 learning and prefer to receive more comprehensive CF. As for CF timing, responses revealed preferences for a shorter time between making errors and receiving CF. Responses also revealed a number of concerns that mediate preferred timings for receiving CF, including concerns about time, interruption of ideas, accuracy of automated CF, accuracy of text, learning from errors, and manageability of CF.

The next section presents and discusses the results of research question 2, which explores L2 learners' perceptions and preferences of different CF strategies. The results include surveys and semi-structured interviews data similar to question 1 in addition to video-stimulated

interviews for focused investigation of L2 learners' perceptions and preferences of automated graduated CF.

4.3. Research Question 2: Perceptions and Preferences of CF Strategies

Research question 2 enquired about how ESL learners perceive four different strategies of automated written CF in terms of desirability, clarity, and usefulness. These CF strategies are identification, direct correction, metalinguistic CF, and graduated CF. Surveys revealed how perceptions of different CF strategies rank in comparison with each other and how different types of perceptions and preferences influence each other. Further, structured and stimulated recall interviews revealed a number of concerns that appear to affect perceptions of different CF strategies. These perceptions, preferences, and concerns are the topic of in-depth discussion in this section.

Discussion of the results will draw a comparison with previous research in order to identify conflicting and similar results and attempt to explain the reasons behind any inconsistency. However, this comparison will be limited by the few studies that quantitatively examined perceptions of CF strategies and the number of comparable CF strategies and types of perceptions examined by them. Specifically, discussion of CF strategy perception results will mainly draw a comparison with the results of Amrhein and Nassaji (2010) and Leki (1991), the only two CF studies found quantitatively to measure perceptions of CF strategies. However, implementation of metalinguistic CF in the two studies is not fully comparable to its implementation in the present study. In the present study, metalinguistic CF explains why a sentence is erroneous and suggests using a specific grammar construction instead; whereas, in Amrhein and Nassaji (2010) and Leki (1991), metalinguistic CF just labels the error type in the sentence. Furthermore, perceptions of desirability and clarity were not examined by these two

studies, preventing comparison of desirability and clarity perceptions across different learner populations and settings. In light of these caveats, perceptions of CF strategies are reported and discussed next.

As shown in Table 12, survey results identified that overall the most desirable CF strategies, listed from most to least desirable, were direct correction ($\mu=4.30$, $sd=1.02$, positive rating = 83%), metalinguistic CF ($\mu=3.73$, $sd=1.20$, positive rating = 73%), identification ($\mu=3.73$, $sd=1.14$, positive rating = 70%) and graduated CF ($\mu=3.53$, $sd=1.28$, positive rating = 60%). Similar strong preferences for direct correction were reported in the literature (Amrhein & Nassaji, 2010; Leki, 1991). This strong preference for direct correction stands in stark contrast to educational practices that favor less direct CF strategies on the basis that less direct CF pushes learners to have a more active role in the error correction process and leads to greater L2 development (Ferris, 2006; Ferris & Roberts, 2001). This preference mismatch between learners and actual educational practices has some researchers and educators concerned about its potential impact on the effectiveness of CF (Ferris, 2011). However, for L2 learners, a number of concerns such as time constraints and goal of writing appear to mediate preference for direct correction. These concerns are discussed in more detail below when reporting interview data related to the perception of direct correction.

Clarity ratings show that direct correction ($\mu=4.77$, $sd=0.43$) is perceived as the clearest CF strategy followed by graduated CF ($\mu=4.60$, $sd=0.62$), metalinguistic CF ($\mu=4.20$, $sd=0.96$), and identification ($\mu=4.03$, $sd=1.19$), respectively. Although differences between mean values are not big, variabilities in responses as indicated by standard deviation values are much bigger in the last two strategies. Small standard deviation indicates stronger agreement between learners

on the clarity of direct correction and graduated CF, and the opposite is true for metalinguistic CF and identification.

Table 12

Summary of Responses to CF Strategies' Perception and Preference Statements on a 5-Point Likert Scale (n=30)

Item	(Strongly) Agree (Very) Useful / Clear	Neutral	(Strongly) Disagree (Very) Unuseful / Unclear	Mean	SD
5. Identification					
5.a. desirability	21 (70%)	3 (10%)	6 (20%)	3.73	1.14
5.b. clarity	22 (73%)	2 (7%)	6 (20%)	4.03	1.19
5.c. usefulness	27 (90%)	0	3 (10%)	4.20	1.00
6. Direct Correction					
6.a. desirability	25 (83%)	3 (10%)	2 (7%)	4.30	1.02
6.b. clarity	30 (100%)	0	0	4.77	0.43
6.c. usefulness	29 (97%)	0	1 (3%)	4.60	0.81
7. Metalinguistic CF					
7.a. desirability	22 (73%)	2 (7%)	6 (20%)	3.73	1.20
7.b. clarity	26 (86%)	2 (7%)	2 (7%)	4.20	0.96
7.c. usefulness	23 (76%)	1 (3%)	6 (20%)	4.03	1.25
8. Graduated CF					
8.a. desirability	18 (60%)	4 (13%)	8 (27%)	3.53	1.28

Table 12
continued

8.b. clarity	28 (93%)	2 (7%)	0	4.60	0.62
8.c. usefulness	24 (80%)	1 (3%)	5 (17%)	4.13	1.28

Overall, based on clarity ratings, participants agreed that direct correction and graduated CF provided the clearest and most easily understood information about form errors. In the follow-up interviews, participants shared a number of concerns that explain these clarity perceptions. Participants were concerned that unfamiliar meta-language in metalinguistic and graduated CF and the lack of specific details about errors in identification could make CF less clear and more difficult to apply correctly. Indeed, this is one of the trade-offs of indirect CF strategies that encourage more active learner participation in the error correction process. Metalinguistic CF and identification provide clues and partial information so that learners rely more on themselves to correct errors. However, if the clues and partial information are not clear enough due to unfamiliar meta-language or insufficient information, learners may not be able to make successful use of indirect CF.

Survey items that examined how useful participants perceive different CF strategies revealed comparable results to clarity perceptions. Namely, direct correction ($\mu=4.60$, $sd=0.81$) was ranked as the most useful CF strategy while other strategies had comparable perceptions: graduated CF ($\mu=4.13$, $sd=1.28$), metalinguistic CF ($\mu=4.03$, $sd=1.25$), and identification ($\mu=4.20$, $sd=1.00$). Close clarity and usefulness perceptions of CF strategies are not surprising knowing that learners are more likely to make successful use of clear and well understood CF than unclear and poorly understood CF. Similar usefulness perceptions of different CF strategies were revealed by Amrhein and Nassaji (2010). Specifically, they found that learners believe the

lack of specific information and unfamiliar meta-language in indirect CF strategies make CF less useful for error correction.

In the following sections, preferences and perceptions of individual CF strategies are reported and discussed in light of survey and interview results. This discussion will examine how learners perceive different CF strategy and how their perceptions and concerns influence their preferences for CF strategies.

4.3.1. Identification

Survey items 5.a., 5.b., and 5.c., respectively, enquired about how desirable, clear, and useful participants perceived identification alone as a CF strategy. In this CF strategy, errors are highlighted with a yellow background with no additional information about the error type or how it should be corrected. Of total responses, 70% indicated desirability of identification while 10% indicated neutral desirability and 20% indicated no desirability at all. The mean of desirability responses ($\mu=3.73$, $sd=1.14$) shows an overall neutral desirability of identification. On the other hand, the means of clarity responses ($\mu=4.03$, $sd=1.19$) and usefulness responses ($\mu=4.20$, $sd=1$) show moderately positive perceptions of the clarity and usefulness of identification.

Despite perceiving identification as a clear and useful CF strategy, learners have a neutral desire for it. This neutral preference for identification despite positive clarity and usefulness perceptions may reflect learners' awareness and past experience with other CF strategies. In fact, factor analysis results reported in Table 9 show that perceptions and preference of identification negatively associates with how frequent AWE tools are used by learners. This suggests that desirability for identification lessens as AWE tools are used more frequently by learners. This connection between perceptions of identification and AWE usage frequency is discussed later in the section about past AWE experience. Furthermore, other perceptions and concerns were

identified in the interviews with potential influence on preferences for identification. These additional perceptions and concerns are reported and discussed later in this section.

Factor analysis results also show high factor loadings of desirability (.80), clarity (.73), and usefulness (.68) of identification with factor 3, *Identification and Frequency of AWE Use*. This shows that desirability of identification is strongly related with how clear and useful it is perceived. Finding a connection between perceptions and preferences is not surprising. In fact, it should be expected that learners' CF preferences reflect how they perceive different CF strategies. Factor 3 shows that preference for identification is closely connected to clarity and usefulness perceptions. This connection is revisited again when reporting metalinguistic CF and direct correction results.

Means comparison shows that the mean of identification usefulness responses ($\mu=4.20$, $sd=1.00$) reported here is greater than the means reported by Amrhein and Nassaji (2010) ($\mu = 2.4$, $sd = NA$) and Leki (1991) ($\mu = 2.22$, $sd = 1.19$). Learner populations across the three studies are somewhat comparable in age, L2 proficiency level, target language, and educational setting (i.e., ESL settings). Therefore, the more favorable perception of identification identified in the present study could only be attributed to differences in CF source: teacher vs AWE tool. It could be argued that the limited time of their teacher may cause learners to believe that identification alone is not useful. In other words, learners may feel that identification alone is not useful for error correction if their writings are only marked once or twice by their teachers. On the other hand, identification could be more appealing to learners in automated settings because there is no limit to how many times their writings are marked. Therefore, learners are allowed to try as much as they want to correct their errors and receive immediate feedback. In this sense, automated error identification could be more useful than error identification by teachers.

Concern about CF accuracy could be an additional factor with potential influence on usefulness perception of identification. In the interviews, a few participants were concerned about the accuracy of automated CF. In contrast, CF literature shows no such concern towards CF from teachers. With this in mind, it could be argued that because learners have high confidence in CF from teachers they may feel that it is more useful for error correction if teachers use other more detailed and informative CF strategies. In contrast, low confidence in the accuracy of automated CF could lead learners to feel that accurate identification of errors is more useful than detailed and informative CF strategies that may be inaccurate. This is a reasonable concern knowing that automated error identification is easier and less prone to errors than more involved automated CF strategies that provide information about the cause of the error and how to possibly correct it.

Follow-up semi-structured interviews revealed concerns about the amount of information provided by identification and a potential mismatch with L2 proficiency level. First, a number of participants ($n=9$) indicated that identification alone is not clear and useful enough for error correction because of the little amount of information it provides for understanding errors as illustrated by the following quotes:

Because only highlighted is not really clear. Maybe this sentence is clear but if I think other sentence, like not only this grammar error there is another mistake in grammar, it would be confusing maybe. (Participant 21, semi-structured interview)

I think I need the software to tell me the correct word choice (Participant 6, semi-structured interview)

The highlights can tell the wrong but I don't know how to correct it, I need Google or other search (Participant 20, semi-structured interview)

Amrhein and Nassaji (2010) have reported a similar concern about identification from their participants. This concern could explain why identification was ranked the least clear CF

strategy on the surveys. Furthermore, this perceived lack of sufficient information in identification had one participant concerned about its suitability for non-advanced learners. This participant stated that error identification alone requires advanced knowledge of different types of grammar errors and how to fix them as illustrated by the following quote:

I think this is good for people who are advanced ESL user, English second language. And for the intermediate or beginner the more progressive way will be more helpful. The beginner they may need more guidance and some opinion. But I think for advanced learners, they have those guidance in their mind so you just have to give them quick remind and they will react to their error (Participant 23, semi-structured interview)

Indeed, identification may not lead to successful error correction if the type and cause of error are unknown to learners. However, this is not entirely true about error identification. A number of participants agree that identification also helps draw learners' attention to errors ($n=4$) as illustrated by the quotes below. Once learners are aware of their errors, they can correct the errors either by themselves or by consulting with external resources.

I think better than having nothing because it's hinting there is something going on. It kinda shows me, "Oh, flashing, there is something." (Participant 1, semi-structured interview)

Normally, everything that is highlighted it draws closer attention. I normally do it when I am reading papers. When I highlight, even if I take a week there is a reason why I am highlighting something... and yellow is a very sharp color. It is like, hey!! Warning here! (Participant 24, semi-structured interview)

To recap the information presented thus far, survey and interview data revealed that identification is the subject of a number of concerns and perceptions that shape how much it is preferred by L2 learners in comparison with other CF strategies. Factor analysis identified an association between clarity and usefulness perceptions and preference for identification. It also identified a negative association between frequency of AWE use and preference for identification suggesting that experience with other CF strategies may raise learners' expectation

for the kind of CF that AWE tools can provide. Finally, the interviews identified three major concerns that may influence preferences for identification. These concerns were related to the accuracy of automated identification, the insufficient amount of information about errors provided in identification, and the potential mismatch between the amount of information provided in identification and the L2 proficiency level of learner.

The next section continues with presenting and discussing survey and interview results. It focuses on learners' preferences and perceptions of direct correction by examining factor analysis results for associations between preference and perception items and interviews for major concerns and perceptions not accounted for in the surveys.

4.3.2. Direct correction

Perceptions of desirability, clarity, and usefulness of direct correction were measured by survey items 6a, 6b, and 6c, respectively. Direct correction involves identifying errors through highlighting and supplying correct language forms and constructions. As mentioned earlier, direct correction ranked the highest on desirability ($\mu=4.30$, $sd=1.02$), clarity ($\mu=4.77$, $sd=0.43$), and usefulness ($\mu=4.60$, $sd=0.81$) in comparison with other CF strategies included in the survey. Participants have highly positive perceptions of direct correction. Two past CF studies (Amrhein & Nassaji, 2010; Leki 1991) reported similar results about usefulness perceptions of direct correction. Specifically, these studies agreed that learners perceive direct correction as more useful than error identification alone. Moreover, Amrhein and Nassaji (2010) found that favorable usefulness perception of direct correction far exceeded perceptions of all other CF strategies included in their study. However, Leki (1991) partially disagrees with Amrhein and Nassaji (2010) by reporting higher usefulness perceptions for some CF strategies that were not included in the present study. This may indicate that learners in some situations or settings may

perceive other CF strategies as more useful than direct correction. The factor analysis and follow-up semi-structured interviews have identified factors that may mediate perceptions of direct correction and explain why it may be viewed less useful in certain situations and settings. These results are discussed next.

Factor analysis revealed two main associations related to direct correction in factor 2, *Direct Correction with Comprehensive CF*. First, factor 2 shows a strong positive association between favorable perceptions of comprehensive CF and direct correction. This association was discussed earlier when comprehensive CF results were reported. Second, factor 2 shows a strong positive association between perceptions of usefulness (.92 loading) and desirability (.71 loading) of direct correction. Basically, this shows that learners who showed preference for direct correction believed it is useful for error correction. Factor analysis has revealed a similar connection between preference and usefulness perception of identification and metalinguistic CF. Obviously, learners' preferences of CF strategies are mediated by usefulness perceptions. However, based on follow-up interviews with participants, other perceptions not examined by the survey may have an influence on learners' preferences of CF strategies as reported next.

Semi-structured interviews revealed a number of perceptions that could be behind learners' preference for direct correction strategy. These perceptions relate to time, complexity/clarity, and learning. First, direct correction appears to address learners' concern about the time needed to identify and successfully correct errors. A number of participants stated that direct correction saves time by not requiring them to identify and correct their errors ($n=7$) and causes less distraction from other writing goals ($n=1$) as illustrated by the following quotes:

Yeah, because when I'm writing like articles or large texts, I need, I need to, I need the error correction, make more professional, to do faster, and save time and saves the... Because I'm also... the main reason I'm not good in spelling and I'm not good in grammar. That's the big problem for me. (Participant 10, semi-structured interview)

I think sometimes I just need I just need them to give me just... because I'm writing and having another stuff in my mind so just give me the right answer and then move on. (Participant 14, semi-structured interview)

Second, some participants believe that direct correction is easy to work with because it is not as complicated as metalinguistic and graduated CF and clearer than identification alone ($n=5$). These participants believe simplicity and clarity of direct correction lead to less confusion and fewer unsuccessful error correction attempts as illustrated by the following quotes:

that one [direct correction] is easier to get the answer. (Participant 15, semi-structured interview)

If it does for you like there is any error and then they give you like the correct form, that one eases your work. It is easier to work with. (Participant 24, semi-structured interview)

Third, some participants voiced their concern about learning from direct correction. These participants believe that direct correction does not help them learn how to correct their errors ($n=6$). However, other participants believe that direct correction is more useful for learning than identification alone ($n=1$) and especially useful when correct answers are unknown ($n=1$). These perceptions are illustrated by the following quotes:

It [direct correction] can be useful but it's not really useful to learn (Participant 21, semi-structured interview)

Yes because it gives you right answer but it does not help you to learn how to get it or why it's the right answer. It helps you to correct word and that's it. (Participant 18, semi-structured interview)

It provides not just the correct answer, sometimes I correct the answer but I don't know why it is wrong. (Participant 4, semi-structured interview)

I think highlight plus correction, you learn more and it saves you more time. You wouldn't want to be always in guessing. (Participant 24, semi-structured interview)

...sometimes we don't know, I don't know what the past tense is, so what the correct word is. (Participant 5, semi-structured interview)

Based on the interviews, preference of direct correction appears to be influenced by time, complexity/clarity, and learning perceptions. Obviously, direct correction should be very appealing to learners who prefer to spend less time correcting errors. By providing overt and direct corrections, this strategy does not require learners to think about what the actual errors are and how to fix them. This is line with past studies that found direct correction leads to much faster error correction process (Baker & Bricker, 2010) and requires less time from learners (Lee, 2005). However, CF is not all about producing error-free text. In many settings, it is also intended to lead to L2 development. Concern about learning from direct correction brings attention to the point that correcting errors without active learner involvement may not lead to actual L2 development. Yet, evident from the strong preference for direct correction in survey data, this concern about learning from direct correction does not substantially impact preference for this CF strategy. One potential explanation for this is that participants in the present study may be at an L2 proficiency level where they believe they do not need to spend much time on correcting surface level errors. They also may simply be giving more weight to immediate time constraints than long-term learning goals.

Overall, data reported here revealed a number of perceptions of direct correction related to usefulness, time, complexity/clarity, and learning that mediate preferences for this strategy. By facilitating error correction and satisfying concerns about time and feedback complexity, direct correction holds greater appeal for most learners. Qualities of direct correction that render it very useful for error correction and that satisfy these concerns should guide implementation and use of CF strategies in AWE tools. Unlike most ESL teachers who are pressed for time, AWE tools

are capable of offering flexible and personalized CF that can be mindful of learners' diverse concerns.

The next section presents and discusses preference and perception data of metalinguistic CF. First, it examines responses to survey items about desirability, clarity, and usefulness of metalinguistic CF. Then, it examines factor analysis results that pertain to metalinguistic CF items on the survey. Finally, it examines interview data that may reveal perceptions unaccounted for in the survey.

4.3.3. Metalinguistic CF

Survey items 7a, 7b, and 7c measured perceptions of desirability, clarity, and usefulness of metalinguistic CF, respectively. In this strategy, metalinguistic explanation is provided next to highlighted grammar errors. Responses to item 7a show neutral desirability ($\mu=3.73$, $sd=1.20$) for metalinguistic CF while responses for items 7b and 7c show positive perceptions of clarity ($\mu=4.20$, $sd=0.96$) and usefulness ($\mu=4.03$, $sd=1.25$), respectively. This result is part of two patterns that can be seen in the survey response. First, clarity and usefulness responses have higher means than desirability responses. This indicates that desirability of CF strategies may be influenced by additional perceptions and concerns not measured by the survey. Some of these perceptions and concerns have been revealed in the interviews. Later in this section, these perceptions and concerns are discussed. Second, higher means of clarity and usefulness responses are accompanied with higher means of desirability responses. For instance, more positive clarity and usefulness perceptions of direct correction are accompanied with more preference for this strategy than metalinguistic CF that has less positive clarity and usefulness perceptions. As presented next, the factor analysis results consolidate this conclusion by

identifying an association between clarity and usefulness perceptions and preferences for metalinguistic CF.

Desirability (.85), clarity (.70) and usefulness (.84) survey items of metalinguistic CF have high loadings with factor 1, *Desire for Clear and Useful CF*. These are the highest loading items for factor 1. The similar loadings of these items indicate that preference for metalinguistic CF is highly related to how clear and useful it is perceived by learners. This adds further support to the presence of a connection between clarity and usefulness perceptions and preference of CF strategy. As mentioned in earlier sections about other CF strategies, the presence of this connection is unsurprising as preferences are expected to be influenced by how learners perceive different CF strategies.

Follow-up semi-structured interviews revealed additional concerns and perceptions that may influence learners' preference for metalinguistic CF. These concerns and perceptions relate to learning, time, and L2 proficiency level. First, many participants believe that metalinguistic CF leads to learning because it provides information about the error ($n=13$), pushes learners to think about and review their errors ($n=1$), and leads learners to avoid making similar errors ($n=1$). These perceptions are illustrated by the following quotes:

It's good to to learning, or to improve your learning about correct and to the type of error and how to correct in... in others writing. (Participant 13, semi-structured interview)

It's good to learn and to understand. (Participant 10, semi-structured interview)

...this can make people think (Participant 25, semi-structured interview)

For the people who are in my level of writing... so I am not the advanced writer that's why I like that type of thing. So it is teaching me how to correct the errors but may be after ten years later I wouldn't prefer may be that approach... may be I will prefer it should correct my error immediately wouldn't focus on my [errors] that much so I like here because teaching me how to... (Participant 25, semi-structured interview)

...because my English is not my first language of course when I finish writing lets say an essay of course I will see some mistakes that I don't know why the computer is saying this is a mistake so that's why for my situation I need [the computer] to explain in detail why this is a mistake. (Participant 26, semi-structured interview)

However, some participants are concerned about the time required to read and apply explanation especially when learners are pressed for time ($n=5$) as illustrated by the following quotes:

But it's not good... and there is time... there is not enough time to learn. I'm just want to finish my work or my text and submit it or e-mail... (Participant 10, semi-structured interview)

I think for my English this is too much information. And if I have more errors than I have more passages like this... I am going to waste more time. (Participant 6, semi-structured interview)

Other interview responses reflect beliefs about the level of learners who may need metalinguistic CF and the type of language that should be used in metalinguistic CF. Specifically, a number of participants think that explanation may be more helpful for beginners who do not know why their text is grammatically incorrect ($n=4$). They believe advanced learners may not need metalinguistic CF because they already know this information. These perceptions are illustrated by the following quotes:

For the people who are in my level of writing... so I am not the advanced writer that's why I like that type of thing. So it is teaching me how to correct the errors but may be after ten years later I wouldn't prefer may be that approach... may be I will prefer it should correct my error immediately wouldn't focus on my [errors] that much so I like here because teaching me how to... (Participant 25, semi-structured interview)

I am now studying English for academic study, for example it will be useful to use this tool if I'm an ESL student but after that I don't think it is necessary to know, to know what type of error I mean if it is present simple or past simple (Participant 12, semi-structured interview)

However, some participants are concerned about the language level used in metalinguistic CF especially for beginner L2 learners. They think that unfamiliar meta-language used in metalinguistic CF could make explanation harder to understand and apply ($n=5$). Some suggested that metalinguistic CF should use simpler language for basic error types ($n=2$). These perceptions are illustrated by the following quotes:

Because I don't know what is simple past and what is present perfect. So I can try but I don't know what the tense is that word. So if I don't know then I won't understand what the correct word or the sentence going to be. (Participant 5, semi-structured interview)

It's too much word to explain. Therefore... I am not native student so maybe half are new words I need to translate or a something. So maybe make me confuse. (Participant 22, semi-structured interview)

Finally, one participant as quoted below pointed to an interesting use of explanation. According to this participant, an explanation can be used to verify if an error detected by software is an actual error and not a case of false positive. This use of explanation appears to address this participant's concern about the accuracy of automated error detection.

I think for me there are some kind... there is some kinds of similarities, that is what I'm saying, because some of the... in Microsoft Word, the problem they don't know the error so they will just give you this sentence. Okay. I think it's useful, okay, I agree with that, but sometimes they will not give you the right... they don't know if there is an error or not. So, yeah, I prefer them to say that there is some error and this is what we think is an error. (Participant 14, semi-structured interview)

Overall, the interviews identified a number of concerns and perceptions that may influence L2 learners' preferences for metalinguistic CF. Individual learners may give different weight to these concerns depending on their L2 proficiency level and writing goal. Many participants agree that metalinguistic CF can help in L2 learning. However, concern about time to complete writing tasks may take more precedence for some L2 learners over long-term learning goals. Survey responses reflected this interaction between concerns. Namely, these

concerns could be causing learners to have a lower preference for metalinguistic CF despite having favorable clarity and usefulness perceptions.

The next section presents and discusses results pertaining to graduated CF. This is the last CF strategy examined by the present study. Being a novel automated CF strategy, the perception of graduated CF is also examined by video-stimulated interviews. Results of video-stimulated interviews in addition to surveys and semi-structured interviews are presented and discussed next.

4.3.4. Graduated CF

Graduated CF is the fourth and last CF strategy examined by the surveys, semi-structured interviews, and video-stimulated interviews. Graduated CF involves initiating an automated dialog with participants for the purpose of leading learners to successfully self-correct their errors with the minimum amount of assistance possible. Survey and interview results show how L2 learners perceive graduated CF, how L2 learners' perceptions affect their preference for graduated CF, and what concerns L2 learners have about graduated CF.

Survey items 8a, 8b, and 8c, respectively, enquired about desirability, clarity, and usefulness of graduated CF. Unlike other CF strategies examined here, responses show a bigger gap between desirability of graduated CF and perceptions of its clarity and usefulness. Responses show that graduated CF is the least desirable CF strategy ($\mu=3.53$, $sd=1.28$) despite ranking the second and third most clear ($\mu=4.60$, $sd=0.62$) and useful ($\mu=4.13$, $sd=1.28$) CF strategy. Obviously, other concerns and perceptions are mediating preference for graduated CF. These concerns are explored in semi-structured and video-stimulated interviews later in this section after discussing factor analysis results pertaining to graduated CF items.

Factor analysis did not identify any association between usefulness and clarity perceptions of graduated CF. This may indicate that clarity perception of graduated CF has no substantial effect on usefulness perception. Other perceptions and qualities of graduated CF may have a greater effect on usefulness perception. Based on the high loading of usefulness item (.59) with factor 1, usefulness perception appears to be mediated by similar qualities that mediate usefulness perception of metalinguistic CF. A key quality shared between metalinguistic CF and graduated CF is explaining errors through meta-language. Based on the interviews, learners appear to associate learning with this quality of metalinguistic CF. Thus, it can be confidently assumed that usefulness perception of graduated CF is associated with usefulness perception of explaining error through meta-language.

Semi-structured interviews revealed two common concerns about graduated CF: concerns about learning and time. Other interview responses reflected mixed feelings about graduated CF. First, interview responses showed that many participants believe graduated CF helps in learning ($n=15$) and causes them to think about and remember correct grammar rules ($n=8$) as illustrated by the following quotes:

If you have no idea about the grammar rules when you use the program it helps you to improve your grammar and know the mistakes. If you are in your hometown you can learn English by this program when it gives you the mistake and the correction and how to write you know the differences between that tenses (Participant 11, semi-structured interview)

Actually i prefer when I get the corrected word, corrected answer and I like the way that he ask me about the correction may be but didn't like the way when he just highlighted the wrong answer so ya Because I think the way that he give you the corrected answer you can learn from that way as well as the other way when he ask you about the grammar tense or ask you about some question regarding your wrong answer so you wanna learn from that. (Participant 17, semi-structured interview)

I mean this is very helpful especially when I cannot determine and I need some help then I get the computer interaction with me to give me some hands and better options to choose from. Yes this guidance will be very helpful to teach the learner and to you can make him or her choose the right word. (Participant 27, semi-structured interview)

because... give you some time to think about what you know you already know about the grammar so you can remember how to say emphasise... What you know, what you learned, so just keep this in your mind next time. (Participant 19, semi-structured interview)

Some attributed learning from graduated CF to the dynamic and interactive approach to explaining errors. They believe that guiding and leading learners to successful error correction by withholding information and prompting them for answers should lead to learning. However, one participant thought that withholding information and prompting learners for answers does not help them learn if they do not already know the answers. This participant may not have realized that information is provided to learners after one or two unsuccessful attempts to ensure that those who do not know or are unable to remember relevant information about the error and how to correct it are introduced to and reminded of such information.

A number of participants voiced concern about the time needed to receive and apply graduated CF ($n=14$) as quoted below:

...in my mind that every time I write... because I write a scientific paper or something I need to done it very fast. (Participant 14, semi-structured interview)

I like the easy way just to show me what's the error and give me the correction option so in short words because if you correct in one paragraph or one error in paragraph it will consume a lot of time and effort. ...I am talking about discussion I would like the correction to be just immediately and in one word. (Participant 28, semi-structured interview)

Because it is interrupting me while writing for example I wrote one sentence and it's asking me two questions so it's a lot of time so I will spend much of time. I will need to optimize the trade-off between learning how to correct and focusing on writing I think it's a trade off. (Participant 25, semi-structured interview)

In case you are targeting that the user to correct only correct mistakes of the user I think this is a waste of time. (Participant 26, semi-structured interview)

These participants believe that graduated CF requires too much time especially for learners who are pressed for time or focusing exclusively on correcting errors instead of learning about errors. In such cases, some believe that graduated CF becomes less desirable ($n=3$). As discussed earlier, similar concerns about time were identified for metalinguistic CF but to a lesser extent than for graduated CF. Concern about time for metalinguistic CF and graduated CF is not surprising knowing that indeed these two CF strategies require more time from learners to process and apply with the latter requiring more time than the former.

Other responses show mixed feelings and concerns about graduated CF. On the positive side, some participants believe that graduated CF promotes learner-autonomy ($n=1$), provides tutorial guidance suitable for beginner L2 learners ($n=3$), and resembles interaction with human teachers ($n=1$). In contrast, others were concerned about the flexibility of graduated CF to accept different but valid learners' responses ($n=1$), complexity of interaction involved in graduated CF ($n=1$), graduated CF requiring too much effort from learners ($n=1$), and unsuitability of graduated CF for low proficiency learners who lack basic meta-language skills and knowledge. These concerns are not incompatible with one another. Indeed, language level, interactional complexity, and L2 proficiency level of learners are important factors that should be considered when designing and integrating graduated CF in AWE tools.

Overall, semi-structured interviews revealed two major competing concerns about graduated CF: time and learning. Participants voiced the same concerns about metalinguistic CF, which reflect similarity between these two CF strategies. Despite seeing a learning value for graduated CF, participants are very concerned about the time required to use this strategy for

error correction successfully. At least for the learner population included in the present study, concern about time takes more precedence over learning concern leading to less preference for graduated CF.

Data from video-stimulated interviews complement semi-structured interviews by offering deeper insights into learners' perceptions of graduated CF. In video-stimulated interviews, participants offered their opinions and suggestions about the major interactional steps and features of graduated CF as implemented and presented by *wTutor*. These interactional steps include highlighting errors of the same type, questioning learners about their errors, prompting learners to select an answer from a list, prompting learners to correct example sentences and, explaining grammar rules with questions. In addition to these interactional steps, participants shared their perceptions of the usefulness of graduated CF, the language level used in the tool, multiple correction attempts, and the L2 proficiency level for which graduated CF could be most suitable. Most of the beliefs shared in the video-stimulated interviews are similar to the beliefs identified in semi-structured interviews as will be shown next.

Questioning learners about their errors is a key component in graduated CF. Indeed, questioning is important for identifying whether an error reflects a lack of knowledge about relevant grammar rules and how to apply them or reflects an unintentional mistake. Subsequent interaction in graduated CF is based on learners' responses to these questions. A number of participants believe that questioning about errors is helpful for learning ($n=9$) as illustrated by the following quotes:

when it was first asked, I kinda, for example, consciously thought whether this is a verb form error or whether this is basically a tense error or yeah it kinda I think pushed me to think about the verb you know the error types and then kinda let me to think about them more proactively what type of errors would they be. I kind of pushed myself to think about those error types. (Participant 1, video-stimulated interview)

The procedure helped yes, narrowed my thinking so it helped (Participant 9, video-stimulated interview)

The way the tool converses is it does not give me the answer immediately. It asks me questions. As if it tries to push me to come up with the answer. Or it tries to help me remember English grammar rules. It's possible I don't know how to correct, don't know there is an error, and unable to remember to correct word. But, I like this questioning strategy. It is gradual in a way that helps in remembering grammar rules and identify the error even before the tool gives me the answer. I find out the correct word before the tool tells me. Simply by following his questions. (Participant 16, video-stimulated interview)

Specifically, they believe that questioning leads learners to recall and apply old information ($n=5$), questioning pushes learners to actively think about their errors ($n=3$), and receiving correction after incorrect responses leads to remembering information better ($n=1$). In terms of learning, these benefits render graduated CF superior to direct correction according to some participants ($n=2$).

Initial questioning is accompanied with highlighting instances of one error type. This highlighting resulted in mixed feeling from some participants. A number of participants believe highlighting all instances of target error type saves time spent in correcting similar errors and focuses learners' attention on those errors ($n=2$) as illustrated by the following quote:

I think it's pretty good because, when we write an article we always use the wrong verb/word sometimes. So, it's very correct to stare at the red, red line, so I can check my answer very quickly...Check your answers, check the answers and see the error (Participant 22, video-stimulated interview)

However, others believe that highlighting does not promote learner autonomy because it does not push learners to find their own errors as illustrated by the following quotes:

sometimes if you just you know this program will tell me what's the mistake and there is highlight you will then don't take care of the writing because its going to give me the mistake because of that. (Participant 11, video-stimulated interview)

If I am beginner who is learning English then I think it is useful when I, if you don't have the [yellow] color on that just tell me that fifteen errors I will just watch, read this sentence carefully. Ok this is past I chose this one choose that one. So for me its concentrate language and once we concentrate the say correct so something like encourage me. If wrong I must know where is it wrong so I say ok, for this one before.
(Participant 15, video-stimulated interview)

Evident here are two opposing concerns: time and learner autonomy. However,, despite the importance of promoting learner autonomy, based on responses reviewed so far, time required to process and apply graduated CF appears to be the primary concern of most participants.

Making list-based responses is among the interactional steps involved in the interaction with *wTutor*. After two or three incorrect attempts to answer some of the questions in graduated CF, participants are prompted to select a response from a list of items. A number of participants showed favorable perceptions of list-based responses ($n=5$). They shared that list-based responses facilitate recall of meta-language, narrows down the number of potential error types, speeds up and facilitates error correction better than typing, and offers an additional chance and alternative approach for correcting errors as illustrated by the following quotes:

Sometimes it's confusing. It's better way to get me the option or get me the correct and ask me is that okay or not. ...for me because I'm not good I'm not good for spelling, to write or something. ...when I write take time then also I'm not good in to write, especially when you use the words 'document' or something because you think it's correct.
(Participant 10, video-stimulated interview)

Learning English is a good way to know that what mistakes and even if we don't know what it is, it shows them because of the multiple choices which gives a good idea to look for the mistakes even if they didn't know by looking at these, "oh yeah, it's past tense".
(Participant 5, video-stimulated interview)

Only one participant voiced concern about a potential mismatch between the meta-language of list-items and the proficiency level of L2 learners. Understandably, concerns about

meta-language have been voiced multiple times in regard to both metalinguistic CF and graduated CF. Issues with unfamiliar meta-language should be anticipated and could be avoided by means of translation or definition as suggested by some participants.

Similar to list-based responses, participants had positive perceptions of being prompted to correct sample sentences. In this interactional step, participants were prompted to correct an erroneous sentence from the original text pre-populating the text editor. Then, they were given immediate evaluation of their correction attempts. Participants believe that correction of sample sentences helps consolidate the information presented in earlier interactional steps ($n=3$), leads learners to remember information better by receiving immediate evaluation after failed correction attempts ($n=1$), and promotes conscious reflection and thinking about how to successfully apply information presented in earlier steps ($n=2$). These perceptions are illustrated by the following quotes:

It is amazing. I mean you will understand how to spell, spelling words or something also it is kind of practicing. You'll try until you achieve it. (Participant 5, video-stimulated interview)

I think this was pretty useful because you need to try that...I need to try, maybe one time, two times to correct the answer, not just show me the answer. Because if the system show me the answer, I will forget very quickly. But like that I need to try, twice or third time, I remember... (Participant 22, video-stimulated interview)

Other comments from participants were concerned about usability issues. Specifically, some participants suggested that more textual context and clearer instructions are needed for this interactional step ($n=2$) and suggested leaving the sample sentence inside the original text to retain context information ($n=2$). Others suggested replacing typing with list-based selection for beginner L2 learners out of concern about the tool's ability to accurately evaluate free, open-ended responses ($n=2$). These perceptions are illustrated by the following quotes:

I think it doesn't have to be popped out, if I can just fix on the paragraph then that would be easier and faster (Participant 21, video-stimulated interview)

Hard to tell I mean, just having one sentence if I know the... what's going on with all the highlighted verbs but if I don't know that's the problems with the highlighted words just having one particular sentence it might not be that helpful for me so... I think have a whole text that would be better than having one popped sentence (Participant 9, video-stimulated interview)

I didn't know what should I write specifically if the software will recognize what I write for it or not. Therefore, it was easy for me to just choose from the list. (Participant 4, video-stimulated interview)

The last major interactional step in graduated CF as implemented by *wTutor* is explanation of grammar rules with questions. In this interactional step, *wTutor* explained the target grammar rule to participants by pointing at relevant parts of speech in a sentence and prompting participants with questions to deduce the error type. This step received favorable perceptions from participants. As shown in Appendix G., participants believe that explaining errors with questions reminds them of correct language use (n=2) and pushes them to think about the cause of errors (n=4) as illustrated by the following quotes:

It's good interaction for the student or for learning (Participant 13, video-stimulated interview)

It's a good strategy because the learner can think in the sentence and ask himself why it asks me... is it not the correct. Because of that he wants to ask himself and think again in this sentence again. (Participant 11, video-stimulated interview)

I think also it's good. It's good way because actually he didn't give you the correct answer directly. He tried to make you speak and explain your opinion about the answer and maybe you have choice maybe to choose the correct answer. I mean maybe have like an idea about the correct answer. So, I think it's good, it's good way. (Participant 17, video-stimulated interview)

Further, they believe explaining with questions focuses a learner's attention on error correction (n=1), promotes learner autonomy (n=2), causes beginner language learners to learn

from their errors (n=1), makes learners confident in receiving enough explanation in order to correct their own errors (n=1). These comments show that participants valued explaining errors with questions for a number of reasons, most of which are addressing their concern about learning from errors. The following quotes illustrate some of these perceptions:

So actually I feel it's very useful also. If you did something wrong they will not just like as they did before they will not just give you the right answer and then you have to know it by yourself. I mean now they will just like give you step by step every time they will try to give you some hints and then you will learn your errors or your mistakes by yourself. I mean overall I think it's much better than just like giving them direct answers.
(Participant 14, video-stimulated interview)

The procedure helped yes, narrowed my thinking so it helped (Participant 9, video-stimulated interview)

...this is helpful; since we have little English and we don't know much so when we try to write something, this is very helpful for us to... we don't need to go to somebody else to correct this. At least most of them can be corrected by itself. So it helps us to correct.
(Participant 5, video-stimulated interview)

Besides perceptions of the interactional steps discussed above, participants have shared some general perceptions and concerns about graduated CF as implemented in *wTutor*. First, many participants believe that interaction with graduated CF is good for learning and practice (n=8) as illustrated by the following quotes:

Ya I think it's a new idea. This is first time I experienced [something like] this. And its very good for those who would like to learn second language and if they are starting its very good for them to have this kind of interaction for them to for the errors to be exact and to have it in their mind. It's much more useful than giving the answer right away.
(Participant 14, video-stimulated interview)

It's good to understand it again so if the computer give me the answer easy come easy go.
(Participant 28, video-stimulated interview)

However, some participants also believe that it may require too much time from learners (n=7) as illustrated by the following quote:

I think this will better better. For example, he found a mistake for me here. I think it's better to tell me directly in order not to waste the time. "This is an error. I think these are the corrections for these errors". Because he is saying oh you have a mistake here, so I start thinking what will be my mistake and I lost my time. I am not sure... It's clear ya but sometimes when I was focussing a lot so I will try to think what's the correction instead of seeing the feedback directly in front of my eyes (Participant 26, video-stimulated interview)

...I think it's not useful tool for like professionals when they try to write like articles or their thesis, dissertation or even assignment because it takes while to interact with the with the software. (Participant 4, video-stimulated interview)

This view echoes similar concerns about time and learning identified in semi-structured interviews. Participants also were concerned about the potential mismatch between meta-language content of graduated CF and the L2 proficiency level of learners ($n=13$) as illustrated by the following quotes:

I think for ESL students for English a second language student, somehow we are not familiar how to describe like simple past tense or like perfect past tense. They have different description. So, in English, somehow we will feel confusing. The terminology to describe the different tense is a little bit unfamiliar with, we are not familiar with the specific tense description. (Participant 23, video-stimulated interview)

I know it... what kind of error but I don't know how to say, how to express in English. But when I see the answers, so I understand each one, so I know now I know how to use how to express in English, simple past. Yeah, I just don't know how to... sometimes I don't know how to express in English. (Participant 19, video-stimulated interview)

I don't think many people will know about determiner, noun, something like that verb. People might know subject verb agreement but determiner the term people might not know about it. (Participant 9, video-stimulated interview)

To bridge this potential mismatch, some suggested accommodating graduated CF for low proficiency levels with simplified language or providing supportive L1 translation or L2 definitions of unfamiliar meta-linguistic terms ($n=21$) as illustrated by the following quotes:

I think this is a great idea but for students, they need to have background information about the grammar. Because if he didn't have for example he will never know the type, present or past. So you need to put in this program some basic information about grammar for like new students, new to grammar. (Participant 12, video-stimulated interview)

I think for the beginner student could be easier for them to use their mother language. Like I am from Taiwan, we use mandarin Chinese. So for the tense description, for the very early beginner, use the mother language will be easier for them. (Participant 23, video-stimulated interview)

I am thinking if you can maybe if you can have them translated to the native language of the learner I mean if you can just put that pointer in there and the translation would appear. (Participant 6, video-stimulated interview)

Further, participants positively perceived the multiple correction attempts allowed during their interaction with *wTutor* ($n=4$). They believe that multiple correction attempts help learners remember information better ($n=2$) while limiting correction to two or three attempts eases their concern about getting stuck during automated interaction ($n=2$). Finally, a number of participants believe that graduated CF is most suited for beginner language learners because it offers opportunities to practice ($n=6$).

Overall, video-stimulated interviews have identified how different interactional steps involved in graduated CF as implemented in *wTutor* address or raise the concerns voiced by participants during semi-structured interviews. First, three interactional steps appear to address learning concern: questioning about errors, prompting to correct sample sentences, and explaining grammar rules with questions. These three interactional steps involve interactively presenting information and encouraging active learner participation and self-reliance. These are the goals of graduated CF and, based on participants' comments, these three interactional steps have met these goals.

On the other hand, time appears to be a concern raised by all interactional steps. This is because each interactional step requires more time from learners regardless of the specific activity involved. In the present study, graduated CF was designed to be less time-consuming and distracting. Specifically, graduated CF was offered only once a specific number of same-type errors was made and detracted once learners showed the ability to self-correct. These two qualities were not obvious to learners because of the limited time they were allowed to use the tool. It is possible that awareness of these time-saving qualities of graduated CF could have mitigated learners' concern about time.

4.4. Research Question 3: Past Experience and CF Preferences and Perceptions

The third and last question of the present study enquired about what connection, if any, exists between learners' past experience with AWE tools and their perceptions and preferences of automated CF and the four CF strategies examined in earlier questions. To answer this question, the results of factor analysis are reviewed to determine the relationship between survey item 9, *How often do you use automated writing evaluation software that show and correct your errors?*, and other survey items. Item 10 collected information about the AWE tools used by participants and their evaluation of these tools. However, item 10 was excluded from exploratory factor analysis due to some missing responses.

For survey item 9, AWE tools were verbally described to participants as software that analyzes text and provides feedback on form, content, and/or organization. Participants were also asked to consider word processing software such as Microsoft Word and LibreOffice Writer as AWE tools because they contain automated analysis and feedback components. As shown in Table 13, responses to survey item 9 show that most participants had very extensive past experience with AWE tools (47% always, 20% often, 23% sometimes, 3% rarely, and 7%

never). This is not surprising given the widespread use of word processors and other specialized writing tools in schools, colleges, and L2 programs within the last two decades.

Table 13

Summary of Responses to CF Experience Statement on a 5-Point Likert Scale (n=30)

Item	Always	Often	Sometimes	Rarely	Never	μ	sd
9. Frequency of past AWE use	14 (47%)	6 (20%)	7 (23%)	1 (3%)	2 (7%)	3.97	1.22

Based on responses to item 10, word processors are the most commonly used AWE tools (78%) followed by *Ginger Page* (7%) and *Grammarly* (3%). *Ginger Page* and *Grammarly* provide form-focused CF and come as both browser add-ons and standalone applications. None of the participants indicated past experience with *MY Access!* or *ETS Criterion*. For each choice of AWE tools selected or added by participants in item 10, participants were prompted to rate their experience with that tool on a five-point Likert scale from *Very Positive* to *Very Negative*. Responses show that most participants have very positive (78%) and positive (13%) experiences with word processors and very few had negative experiences (9%). Similarly, participants who indicated past use of *Ginger Page* and *Grammarly* have positive ratings of their experience with these tools.

Overall, based on items 9 and 10, most participants had positive and extensive experience with AWE tools. According to the cut-off point proposed by Kline (1994), factor analysis results show that item 9 has a moderately high negative loading (-.37) with factor 3. Factor 3 was called *Identification and Frequency of AWE Use* due to the high loadings of the desirability (.80), clarity (.73), and usefulness (.68) identification items. The negative loading of item 9 with factor 3 indicates that preference and positive perceptions of identification have a negative association

with the frequency of using AWE tools. In other words, as the use of AWE tools increases, participants are less likely to prefer or favorably perceive identification. However, factor analysis does not show a positive association between frequency of AWE use and any other CF strategy. This indicates that at least preference and perceptions of other CF strategies is not affected by the frequency of AWE use.

4. 5. Chapter Summary

This chapter presented and discussed the findings of surveys, semi-structured, and video-stimulated interviews addressing the three research questions investigated by the present study. The first research question enquired about perceptions and preferences of automated CF usefulness, frequency, and timing. Surveys and semi-structured interviews revealed overwhelmingly favorable perceptions of automated CF and preference for more comprehensive CF. Furthermore, factor analysis revealed a preference for direct correction with comprehensive CF. Based on the interviews, this preference was explained by the potential effect of direct correction to mitigate concern about the time required to successfully process and apply comprehensive CF. As for timing preferences, the surveys revealed a general preference for receiving CF immediately or shortly after making errors. Participants explained this timing should help them manage error correction more effectively and see immediate improvements in writing accuracy.

The second question enquired about preferences for receiving four different CF strategies and perceptions of the clarity and usefulness of these four strategies. Surveys revealed a stronger preference for direct correction followed by metalinguistic CF, identification, and graduated CF, respectively. Semi-structured and video-stimulated interviews revealed a number of concerns that may mediate perceptions and preferences of different CF strategies. Two major concerns

were time required to correct errors and ability to learn from CF. It was concluded that the time concern took more precedence for the learner population included in the study. This led to a greater preference for direct correction, which requires less time for successful error correction, than metalinguistic and graduated CF, which help in learning but requires more time.

The third research question enquired about the relationship between the frequency of past AWE use and perceptions and preferences of CF strategies. Survey data showed a positive and extensive use of AWE tools by the majority of participants. Further, exploratory factor analysis of survey data revealed a negative association between frequency of past AWE use and preference for identification alone as CF strategy. Accordingly, it was concluded that learners with extensive AWE experience tend to have less favorable perception and preference for identification.

CHAPTER 5. CONCLUSION

The present study explored L2 learners' perceptions and preferences of automated written CF. Specifically, first, this study examined how useful L2 learners perceive automated CF and what frequency and timing they prefer for automated CF. Second, this study examined and compared how desirable, clear, and useful L2 learners perceive four different CF strategies: identification alone, direct correction, metalinguistic CF, and graduated CF. Third, this study investigated what relationships if any exist between frequency of past AWE use and preferences and perceptions of automated CF and its four strategies.

To address these three questions, the expansion and initiation models of mixed-method design were employed. According to this mixed-method design, surveys and semi-structured and video-stimulated interviews were collected, analyzed, and interpreted. Exploratory factor analysis was conducted to determine underlying factors affecting perceptions and preferences of automated CF shared in surveys. Grounded theory approach was employed for coding and identifying recurring themes in semi-structured and video-stimulated interviews. All three data sources contributed in depicting a more complete picture of learners' perceptions and preferences of automated CF.

This chapter summarizes the major findings of the present study according to the three main research questions. Then, it discusses important implications for these findings and shares some recommendations for future research. Finally, it discusses limitations that should be considered while interpreting and applying the findings.

5.1. Summary of Major Findings

Three main research questions were addressed by the present study. The first question enquired about L2 learners' perceptions of automated written CF and their preferences of CF frequency and timing. Corroborating the findings of past research, survey data showed an overwhelmingly positive perception of automated CF. Further, semi-structured interview data showed that participants positively perceive automated CF because it addressed their concerns about writing accuracy and learning. Interview data also revealed concerns about the time needed to process and apply CF, learner-autonomy, and accuracy of automated error detection. These concerns may mediate the desirability and use of automated CF.

In line with past research, survey items that explored preferences of CF frequency showed that most participants prefer comprehensive CF. Based on interview data, this preference is due to participants' desire to have a more accurate text. However, interview data also suggested that preference of comprehensive CF may be mediated by concerns over interruption of ideas, learner-autonomy, and accuracy of automated CF. Factor analysis results indicate that direct correction is preferred with comprehensive CF. Finally, survey data showed that most participants prefer to receive CF shortly after an error is made. Among a wide range of concerns, concern over interruption of ideas and manageability of CF were among the most commonly cited in the interviews about different CF timings.

The second research question enquired about how participants perceive four different CF strategies. Survey data showed that direct correction is the most preferred CF strategy followed by metalinguistic CF, identification, and graduated CF, respectively. Further, factor analysis showed that participants tend to prefer CF strategies that they perceive as more clear and useful. Finally, semi-structured and video-stimulated interview data showed that participants were

mostly concerned about the value of different CF strategies for learning and the time required to process and apply CF. Based on interview data, graduated and metalinguistic CF address participants' concerns about learning because they provide important information about errors and push learners to think about and self-correct their errors. Graduated CF proved appealing for the purpose of learning from CF underscoring the value of adopting a sociocultural perspective to the delivery of automated CF. Providing conversational and dynamic feedback that is responsive to learners' needs and proficiency level is seen by participants as effective for learning and promoting learner autonomy. Maximizing learning and promoting learner autonomy are two of the key objectives of socioculturally inspired CF. On the other hand, direct correction addresses participants' concerns about the time required to process and apply CF. Therefore, it was the most preferred CF strategy, especially with comprehensive CF.

The third research question enquired about the relationship between the frequency of using AWE tools and the perceptions and preferences of automated CF and different CF strategies. Survey data revealed that most participants have extensive past AWE experience mostly through interaction with CF in word processors. Survey data also revealed that most participants have positive perceptions of the AWE tools that they used in the past. Finally, factor analysis results showed that favorable perceptions of identification alone as a CF strategy negatively associate with the frequency of AWE use. This negative association suggests that participants who have extensively used AWE tools expect more than identification alone for CF.

5.2. Implications and Recommendations

The findings summarized above carry significant implications for development and use of automated CF. The findings suggest that developers may need to balance between different CF qualities in order to satisfy the diverse concerns of L2 learners. Further, the findings suggest

that L2 educators may need to actively encourage L2 learners to opt-in for metalinguistic CF and graduated CF whenever possible due to their pedagogical values. This section discusses the implications of key findings pertaining to CF timing, frequency, and strategies in addition to concerns over time and learning.

First, findings showed that participants prefer to receive more frequent and comprehensive CF. Participants link this preference to their belief that frequent and comprehensive CF assist in producing error-free text and enhance their knowledge about grammar errors and how to fix them. Interestingly, findings showed that preference for more frequent and comprehensive CF positively associates with preference for direct corrections. And, according to interview data, preference for direct correction is linked to participants' concern about time required to process and apply CF. This suggests that participants' concern over the time required to process and apply frequent and comprehensive CF is mitigated by direct correction. For developers of AWE tools, this means that they need to balance between the frequency and comprehensiveness of CF and the directness of CF strategies.

Second, findings showed that two main concerns influence participants' preference of CF strategies: time and learning. According to interview data, metalinguistic CF and graduated CF satisfy learners' concern about learning from CF while direct correction satisfies their concern about the time required to process and apply CF. However, preference for direct correction was greater than preference for metalinguistic and graduated CF. This suggests that preference of CF strategies is mostly shaped by time concerns. However, as suggested by some participants, level of language proficiency and purpose of writing may also play a role in whether time or learning concerns have greater influence over preference of CF strategies.

For developers, this means that a “one size fits all” approach to delivering CF may not work. In order to maximize benefits from CF, developers should be mindful of learners’ concerns and how these concerns shape learners’ preferences of CF strategies. Technically, this may require basing the choice of CF strategy on learners’ proficiency level, writing purpose, and other learner factors. Such information can be automatically predicted from learners’ current and previous writings or obtained directly from learners. Otherwise, learners may be better off selecting the type of CF strategy they prefer at the beginning of each writing assignment. On the other hand, L2 educators may also need to point out the value of different CF strategies to learners. This would help learners make informed decisions when choosing between CF strategies.

5.3. Limitations of the Study

The wide range of concerns and their effects on learners’ preferences show that CF should be dynamic and responsive. Graduated CF is intended to be dynamic and responsive to learners’ needs, proficiency levels, and concerns. However, in the present study, graduated CF was ranked the least preferred CF strategy. These seemingly contradictory findings could be explained by the short time given to participants to try and form an informed opinion about graduated CF. The short time may not have allowed participants to fully understand how dynamic and responsive graduated CF is. For example, graduated CF was implemented in *wTutor* such that it only gets triggered once certain number of errors of the same error type are made and prioritized according to the severity of the error type in comparison with other error types in *wTutor* database. This means that, in actual writing tasks, participants would not be prompted with lengthy dialogs with every single error, somewhat satisfying participants’ concern about the time needed to process and apply graduated CF. Also, participants’ L2 proficiency

level may have contributed to the unfavorable perception of graduated CF. Most participants were at an intermediate to advanced proficiency level. However, the language construction targeted for modeling graduated CF was the ESL simple past, which should be quite familiar to all of them. Participants were asked to pretend not know this linguistic construction. This mismatch between proficiency level and target language construction may have undermined the value of graduated CF for participants.

In light of this limitation, future studies investigating learners' perceptions of graduated CF should allow greater amount of time for participants to experience graduated CF. Further, they should ensure matching the level of target constructions with the level of participants' L2 proficiency. They can simulate more genuine interaction between participants and tools delivering graduated CF while allowing greater experience time by having participants complete a composition task that requires them to use a language construction with which they are struggling.

Lack of counterbalancing is another limitation of the present study. Counterbalancing is an approach to design that ensures the order of treatments or measures does not inadvertently affect the outcomes of the study (Cozby, 2009). Ideally, the present study would have counterbalanced the order of the three measures: surveys, semi-structured, and video-stimulated interviews. However, this was not feasible due to the small number of participants.

Finally, user interface is an important factor, which has not been addressed by the present study, and could have a significant effect on CF preferences and perceptions. Some participants have indicated in the interviews that user interface elements such as the size of the avatar and text could affect their preferences. Investigating the effect of user interface on L2 learners'

perceptions and preferences of CF should yield valuable insights and feed into designing and developing more appealing strategies for delivering automated CF.

5.4. Conclusion

This study attempted to fill an important gap in CALL literature pertaining to L2 learners' perceptions and preferences of automated CF in AWE settings. Further, it introduced automated graduated CF as a new dynamic and responsive automated CF strategy. By employing a mixed-method design, the present study tapped into L2 learners' perceptions and preferences and revealed many important concerns. The findings discovered here should guide future development of AWE tools and should provide important considerations for further research.

REFERENCES

- AbuSeileek, A., & Abualsha'r, A. (2014). Using peer computer-mediated corrective feedback to support EFL learners' writing. *Language Learning & Technology* 18(1), 76–95.
Retrieved from <http://llt.msu.edu/issues/february2014/abuseileekabualshar.pdf>.
- AbuSeileek, A. F. (2013). Using track changes and word processor to provide corrective feedback to learners in writing. *Journal of Computer Assisted Learning*, 29, 319–333.
- Agudo, J.D.D.M. (2012). Investigating Spanish EFL students' beliefs and preferences regarding the effectiveness of corrective feedback. *International Journal of Humanities and Social Science*, 2(19), 121-131.
- Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal*, 78(4), 465-483.
- Allwright, D., & Bailey, K.M. (1991). *Focus on the language classroom: An introduction to classroom research for language teachers*. Cambridge: Cambridge University Press.
- Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts and L2 learning. *Studies in Second Language Acquisition*, 28, 543–574. doi:10.1017/ S0272263106060268
- Amrhein, H. R., & Nassaji, H. (2010). Written corrective feedback: What do students and teachers prefer and why? *Canadian Journal of Applied Linguistics*, 13, 95-127.
- American Psychological Association (2001). *Publication manual of the American Psychological Association*. Washington, DC: American Psychological Association.
- Baker, W., & Bricker, R. H. (2010). The effects of direct and indirect speech acts on native English and ESL speakers' perception of teacher written feedback. *System*, 38(1), 75-84.

- Belz, J. A., & Kinginger, C. (2002). The cross-linguistic development of address form use in telecollaborative language study: Two case studies. *Canadian Modern Language Review/La revue canadienne des langues vivantes*, 59(2), 189–214.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17(2), 102–118.
- Bitchener, J., & Knoch, U. (2008). The value of a focused approach to written corrective feedback. *ELT Journal*, 63, 204–211. doi:10.1093/elt/ccn043
- Bitchener, J., & Knoch, U. (2009). The relative effectiveness of different types of direct written corrective feedback. *System*, 37(2), 322–329.
- Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19, 207–217. doi:10.1016/j.jslw.2010.10.002
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. West Sussex, UK: Wiley.
- Brown, A. (2009). Students and teachers' perceptions of effective foreign language teaching: A comparison of ideals. *Modern Language Journal*, 93, 46–60. doi:10.1111/j.1540-4781.2009.00827.x
- Calvé, P. (1992). Corriger ou ne pas corriger. *Canadian Modern Language Review*, 48, 458–471.
- Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, 15, 357 - 366.

- Carroll, S., Roberge, Y., & Swain, M. (1992). The role of feedback in adult second language acquisition: Error correction and morphological generalizations. *Applied Psycholinguistics*, 13, 173-198.
- Cerny, C. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research*, 12(1), 43-47.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage Publications Ltd.
- Chaudron, C. (1988). *Second language classrooms*. New York: Cambridge University Press.
- Chen, C.F., & Cheng, W.Y. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12, 94-112.
- Chukharev-Hudilainen, E. (n. d.). Analyzer. Retrieved from <https://cywrite.engl.iastate.edu/wp/analyzer/>
- Chukharev-Hudilainen, E., & Saricaoglu, A. (2014). Causal discourse analyzer: improving automated feedback on academic ESL writing. *Computer Assisted Language Learning*, 1-23. <http://doi.org/10.1080/09588221.2014.991795>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Colpitts, B.D. F. (2016). Japanese Students' Perceptions of Peer Corrective Feedback in an EFL Classroom. *Acta Humanistica et Scientifica Universitatis Sangio Kyotiensis*, 49, 345-358.
- Cozby, P. C. (2009). *Methods of behavioral besearch: Tenth edition*. New York, NY: McGraw-Hill.

- Cuestas Verjano, A. (2013). Using Skype in a primary class: A case study. *Bellaterra Journal of Teaching & Learning Language & Literature*, 6(2), 49-68.
- Diab, R. L. (2005). Teachers' and students' beliefs about responding to ESL writing: A case study. *TESL Canada Journal*, 23, 28-43.
- Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding Its origins and scope, and preventing harm. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 9–33). West Sussex, England: Wiley.
- Donato, R. (2000). Sociocultural contributions to understanding the foreign and second language classroom. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 27–50). New York: Oxford University Press.
- Donato, R. & McCormick, D. (1994). A sociocultural perspective on language learning strategies: The role of mediation. *The Modern Language Journal*, 78(4), p. 453-464.
- Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 206–25). Cambridge: Cambridge University Press.
- Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–261). Cambridge: Cambridge University Press.
- Dulay, H., & Burt, M. (1974). A new perspective on the creative construction process in child second language acquisition 1. *Language Learning*, 24(2), 253–278.
<http://doi.org/10.1111/j.1467-1770.1974.tb00507.x>
- Egi, T. (2010). Uptake, modified output, and learner perceptions of recasts: learner responses as language awareness. *Modern Language Journal*, 94, 1-21.

- Ellis, N. C. (2007a). Implicit and explicit knowledge about language. In J. Cenoz & N. H. Hornberger (Eds.) *Encyclopedia of Language and Education, Second Edition, Volume 6: Knowledge about Language* (pp. 119-132). Springer.
- Ellis, N. C. (2011). Implicit and explicit SLA and their interface. In C. Sanz, & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 35-47). Washington, DC: Georgetown University Press.
- Ellis, R. (2007b). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 339-360). Oxford: Oxford University Press.
- Ellis, R. (2009a). A typology of written corrective feedback types. *ELT Journal*, 63(2), 97-107.
- Ellis, R. (2009b). Corrective feedback and teacher development. *L2 Journal*, 1, 3-18.
- Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning*, 51(2), 281-318.
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies In Second Language Acquisition*, 28(2), 339-368.
- Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36, 353-371.
- Elwood, J. A., & Bode, J. (2014). Student preferences vis-à-vis teacher feedback in university EFL writing classes in Japan. *System*, 42, 333-343. <http://dx.doi.org/10.1016/j.system.2013.12.023>
- Erlam, R., Ellis, R., & Batstone, R. (2013). Oral corrective feedback on L2 writing: Two approaches compared. *System*, 41, 257-268.

- Faqeih, H. I. (2012). *The effectiveness of error correction during oral interaction: Experimental studies with English L2 learners in the United Kingdom and Saudi Arabia* (Unpublished doctoral dissertation). University of York: England.
- Farrokhi, F., & Sattarpour, S. (2012). The effects of direct written corrective feedback on improvement of grammatical accuracy of high-proficiency L2 learners. *World Journal of Education*, 2, 49–57.
- Feng, H. H., Saricaoglu, A., & Chukharev-Hudilainen, E. (2016). Automated error detection for developing grammar proficiency of ESL learners. *CALICO Journal*, 33(1), 49–70. DOI: 10.1558/cj.v33i1.26507.
- Ferreira, A. (2006). An experimental study of effective feedback strategies for intelligent tutorial systems for foreign language. In J. S. Sichman, H. Coelho, & S. O. Rezende (Eds.), *Advances in Artificial Intelligence - IBERAMIA-SBIA 2006* (pp. 27–36). Heidelberg, Germany: Springer.
- Ferris, D. R., & Roberts, B. J. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10, 161–184.
- Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8, 1–10 .
- Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81–104). Cambridge, UK: Cambridge University Press.
- Ferris, D. R. (2011). *Treatment of error in second language student writing* (2nd ed.). Ann Arbor: University of Michigan Press.

- Finegan, E. (2012). *Language: Its Structure and Use, sixth edition*. Boston: Wadsworth.
- Frear, D., & Chiu, Y. (2015). The effect of focused and unfocused indirect written corrective feedback on EFL learners' accuracy in new pieces of writing. *System*, 53, 24-34.
- Green, J. (1993). Student attitudes toward communicative and non-communicative activities: Do enjoyment and effectiveness go together? *The Modern Language Journal*, 77, 1-10.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274.
- Han, Z. (2001). Fine-tuning corrective feedback. *Foreign Language Annals*, 34(6), 582-599.
- Hartshorn, K. J., Evans, N. W., Merrill, P. F., Richard, F. M., Sudweeks, R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL*, 44(1), 84-109.
- Heffernan, N., Otoshi, J., & Kaneko, Y. (2014). Written feedback in Japanese EFL classrooms: A focus on content and organization. *The Journal of Language Teaching and Learning*, 4(1), 55-68.
- Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL*, 16(2), 416-431.
- Horwitz, E. K. (1990). Attending to the affective domain in the foreign language classroom. In S. Magnan (Ed.), *Shifting the instructional focus to the learner* (pp. 15-33). Middlebury, VT: Northeast Conference on the Teaching of Foreign Languages.
- Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System*, 31, 217-230.
- Iglewicz, B., & Hoaglin, D. (1993). *How to detect and handle outliers*. Milwaukee, Wisconsin: ASQC Quality Press.

- Jeon, M., & Kang, I. (2005). Investigating student preferences in error correction in Korean-language teaching. *American Association of Teachers of Korean*, 5(10), 19-49.
- Jiang, L., & Xiao, H. (2014). The efficacy of written corrective feedback and language analytic ability on chinese learners' explicit and implicit knowledge of English articles. *English Language Teaching*, 7(10), 22-34. doi:10.5539/elt.v7n10p22
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Kaiser, H. (1974). An index of factor simplicity. *Psychometrika*, 39, 31–36.
- Kang, E., & Han, Z. (2015). The Efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal*, 99(1), 1-18. DOI: 10.1111/modl.12189
- Kartchava, E., & Ammar, A. (2014a). Learners' Beliefs as Mediators of What Is Noticed and Learned in the Language Classroom. *TESOL Quarterly*, 48(1), 86-109. doi: 10.1002/tesq.101
- Kartchava, E., & Ammar, A. (2014b). The noticeability and effectiveness of corrective feedback in relation to target type. *Language Teaching Research*, 18(4), 428-452.
- Keck, C., Iberri-Shea, G., Tracy-Ventura, N., & Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and acquisition. In J. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 91–131). Amsterdam: Benjamins.
- Kern, R. (1995). Students' and teachers' beliefs about language learning. *Foreign Language Annals*, 28, 71–92. doi:10.1111/j.1944-9720.1995.tb00770.x

- Kern, R., & Warschauer, M. (2000). Theory and practice of network-based language teaching. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 1-19). New York: Cambridge University Press.
- Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.
- Kozulin, A. (1991). *Vygotsky's Psychology, a biography of ideas*. Cambridge: Harvard University Press.
- Krashen, S. D., & Terrell, T. D. (1983). *The natural approach: Language acquisition in the classroom*. Hayward, CA: Alemany Press.
- Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.
- Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications*. London: Longman.
- Kregar, S. (2011). The relative effectiveness of corrective feedback types in computer-assisted language learning. *Electronic Theses, Treatises and Dissertations*. Paper 2871.
- Lantolf, J. P., & Thorne, S. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.
- Lave, J. & Wenger (1991). *Situated Learning: Legitimate Peripheral Participation*. New York, NY: Cambridge University Press.
- Lee, I. (2005). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing*, 17, 144–164 doi:10.1016/j.jslw.2007.12.001
- Lee, E. J. (2013). Corrective feedback preferences and learner repair among advanced ESL students. *System*, 41, 217-230.
- Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition*, 25, 35–63. doi:10.1017/S0272263103000020

- Leki, I. (1999). Techniques for reducing second language writing anxiety. In D. J. Young (Ed.), *Affect in foreign language and second language learning* (pp. 64-88). Boston, MA: McGraw-Hill.
- Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition*, 26(1), 33-57.
- Li, S. (2010a). *Corrective feedback in perspective: The interface between feedback type, proficiency, the choice of target structure, and learners' individual differences in working memory and language analytic ability*. Unpublished doctoral dissertation. Michigan State University, East Lansing, Michigan.
- Li, S. (2010b). The Effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309–365.
- Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, 97(3), 634-654.
- Lim, H., & Kahng, J. (2012). Review of Criterion®. *Language Learning & Technology*, 16(2), 38-45.
- Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE Publications.
- Loewen, S. (2005). Incidental focus on form and second language learning. *Studies in Second Language Acquisition*, 27, 361–386.
- Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: an experimental study. *Computer Assisted Language Learning*, 19(1), 1-14.

- Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective feedback on L2 knowledge. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 361–378). Oxford, UK: Oxford University Press.
- Long, M. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: multiple perspectives*. New York: Newbury House.
- Long, M. (1996). The role of linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.
- Long, M., & Robinson, P. (1998). Theory, research and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.
- Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal*, 82, 357–371.
- Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, 26, 399–432.
- Lyster, R., & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning*, 59, 453–498.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 19, 37–66.
- Lyster, R. & Saito, K. (2010). Oral feedback in classroom SLA: a meta-analysis. *Studies in Second Language Acquisition*, 32, 265–302.

- Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition*, 28, 269–300.
- Mackey, A. (2006) Feedback, noticing and instructed second language learning. *Applied Linguistics*, 27(3), 405-430.
- Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in SLA: A collection of empirical studies* (pp. 408–452). New York: Oxford University Press.
- Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal*, 82, 338–356.
- Makino, T. (1980). Acquisition order of English morphemes by Japanese secondary school students. *Journal of Hokkaido University of Education*, 30, 101-148.
- Maleki, A. & Eslami, E. (2013). The effects of written corrective feedback techniques on EFL students' control over grammatical construction of their written English. *Theory and Practice in Language Studies*, 3(7), 1250-1257.
- McCargar, D. (1993). Teacher and student role expectations: Cross-cultural differences and implications. *The Modern Language Journal*, 77, 192-207.
- Monteiro, K. (2014). An experimental study of corrective feedback during video-conferencing. *Language Learning & Technology*, 18(3), 56–79. Retrieved from <http://llt.msu.edu/issues/october2014/monteiro.pdf>
- Montgomery, J., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing*, 16, 82-99.
- Mowrer, O. (1960). *Learning theory and the symbolic processes*. New York: Wiley.

- Mubarak, M. (2013). *Corrective feedback in L2 writing: A study of practices and effectiveness in the Bahrain context* (Unpublished doctoral dissertation). University of Sheffield: United Kingdom.
- Muranoi, H. (2000). Focus on form through interaction enhancement: integrating formal instruction into a communicative task in EFL classrooms. *Language Learning*, 50(4), 617–673.
- Nagata, N. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal*, 77, 330–339.
- Nassaji, H., & Swain, M. (2000). A Vygotskian perspective on corrective feedback in L2: The Effect of Random versus Negotiated Help On the learning of English Articles. *Journal of Language Awareness*, 9(1), 34-49.
- Netten, J. (1991). Towards a more language oriented second language classroom. In L. Malav & G. Duquette (Eds.), *Language, culture and cognition* (pp. 284–304). Clevedon, UK: Multilingual Matters.
- Newman, D., Griffin, P., & Cole, M. (1989). *The construction zone: Working for cognitive change in school*. Cambridge: Cambridge University Press.
- Noels, K. A. (2001). Learning Spanish as a second language: Learners' orientations and perceptions of their teachers' communication style. *Language Learning*, 51, 107-144.
- Noels, K. A., Clément, R., & Pelletier, L. G. (1999). Perceptions of teachers' communicative style and students' intrinsic and extrinsic motivation. *The Modern Language Journal*, 83, 23-33.
- Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417–528.

- Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, 36, 573–595.
- Park, H.S. (2010). *Teachers' and learners' preferences for error correction*, Unpublished MA Thesis, California State University, Sacramento.
- Parreno, A. (2013). The effectiveness of the three common types of written corrective feedback in improving grammatical accuracy. *New English Teacher*, 7(1), 102-121.
- Penning de Vries, B. (2015). *Computerised speaking or active: The role of automatic corrective feedback in learning L2 grammar*. JK Utrecht, The Netherlands: LOT.
- Peterson, M. (2012). Learner interaction in a massively multiplayer online role playing game (MMORPG): A sociocultural discourse analysis. *ReCALL*, 24(3), 361-380.
- Plonsky, L., & Mills, S.V. (2006). An exploratory study of differing perceptions of error correction between a teacher and students: Bridging the gap. *Northern Arizona University Applied Language Learning*, 16, 55–77.
- Profozic, N. M. (2013). *The Effectiveness of corrective feedback and the role of individual differences in language learning*. Peter Lang: Bern, Switzerland.
- Rahimi, M., & Zhang, L. J. (2013). The role of incidental unfocused prompts and recasts in improving English as a foreign language learners' accuracy. *The Language Learning Journal*, DOI: 10.1080/09571736.2013.858368.
- Rassaei, E., & Moinsadeh, A. (2011). Investigating the effects of three types of corrective feedback on the acquisition of English wh-question forms by Iranian EFL learners. *English Language Teaching*, 4(2), 97-106.
- Rassaei, E. (2013). Corrective feedback, learners' perceptions, and L2 development. *System*, 41, 472–483.

- Rassaei, E. (2014). Scaffolded feedback, recasts, and L2 development: A sociocultural perspective. *The Modern Language Journal*, 98(1), 417-431.
- Rassaei E. (2015). Oral corrective feedback, foreign language anxiety and L2 development. *System*, 49, 98-109.
- Rosa, E. M., & Leow, R. P. (2004). Computerized task-based exposure, explicitness, type of feedback, and Spanish L2 development. *The Modern Language Journal*, 88(2), 192-216.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research (Vol. 6)*. Thousand Oaks, CA: SAGE.
- Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for second language acquisition: A meta-analysis of the research. In J. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 131–164). Amsterdam: Benjamins.
- Rutherford, W., & Sharwood Smith, M. (1985). Consciousness-raising and universal grammar. *Applied Linguistics*, 6(3), 274-282.
- Sachs , R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition*, 29(1), 67–100.
- Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 229–248). New York: Oxford University Press.
- Santos, M., López-Serrano, S., & Manchón, R. M. (2010). The differential effect of two types of direct written corrective feedback on noticing and uptake: Reformulation vs. error correction. *International Journal of English Studies*, 10(1), 131-154.

- Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning*, 54, 35–78.
- Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning and Technology*, 13, 96-120.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129–158.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.
- Schulz, R. A. (1996). Focus on form in the foreign language classroom: Students' and teachers' views on error correction and the role of grammar. *Foreign Language Annals*, 29(3), 343-364. <http://dx.doi.org/10.1111/j.1944-9720.1996.tb01247.x>
- Schulz, R. A. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corrective feedback: USA-Columbia. *Modern Language Journal*, 85(2), 244-258. <http://dx.doi.org/10.1111/0026-7902.00107>
- Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8, 263-300.
- Sheen, Y. (2007a). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41(2), 255–283.
- Sheen, Y. (2007b). The effects of corrective feedback, language aptitude and learner attitudes on the acquisition of English articles. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: a collection of empirical studies* (pp. 301–322). New York: Oxford University Press.

- Sheen, Y. (2008). Recasts, language anxiety, modified output, and L2 learning. *Language Learning*, 59(4), 835-874.
- Sheen, Y. (2010a). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition*, 32(2), 203–234.
- Sheen, Y. (2010b). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition*, 32(2), 203–234.
- Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, 37, 556–569.
- Shetzer, H., & Warschauer, M. (2000). An electronic literacy approach to network-based language learning. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: concepts and practice*. Cambridge: Cambridge University Press, 171–185.
- Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, 22, 286-306.
- Spada, N., & Lightbown, P. M. (1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition*, 15(2), 205- 224.
- Stefanou, C., & Révész, A. (2015). Direct written corrective feedback, learner differences, and the acquisition of second language article use for generic and specific plural reference. *The Modern Language Journal*, 99(2), 263-282.
- Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing: Case studies. *Studies in Second Language Acquisition* 32(2), 303-334.

- Suzuki, M., (2004). Corrective feedback and learner uptake in adult ESL classrooms. *Columbia University Working Papers in TESOL & Applied Linguistics*, 4, 1-21.
- Swain, M., & Herron, C. (1987). The immersion observation study. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of bilingual proficiency final report: Classroom treatment* (Vol. 2, pp. 190-342). Toronto: Modern Language Centre, The Ontario Institute for Studies in Education.
- Takashima, H., & Ellis, R. (1999). Output enhancement and the acquisition of the past tense. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 173 – 188). Amsterdam: Benjamins.
- Thorne, S. L. (2003). Artifacts and cultures-of-use in intercultural communication. *Language Learning & Technology*, 7(2), 38-67.
- Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytic ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: a collection of empirical studies* (pp. 171–195). Oxford: Oxford University Press.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46, 327–369.
- Truscott, J. (2007). The effect of error correction on learners’ ability to write accurately. *Journal of Second Language Writing*, 16, 255–272.
- Tse, L. (2000). Student perceptions of foreign language study: A qualitative analysis of foreign language autobiographies. *The Modern Language Journal*, 84, 69-84.

- UNESCO (2003). Vygotsky's sociocultural theory. *Education*. Retrieved from http://portal.unesco.org/education/en/ev.php-URL_ID=26925&URL_DO=DO_TOPIC&URL_SECTION=201.html
- Vaezi, S., & Abbaspour, E. (2015). Asynchronous online peer written corrective feedback: effects and affects. In M. Rahimi (Ed.), *Handbook of research on individual differences in computer-assisted language learning* (pp. 271-297). Hershey, Pennsylvania: IGI Global.
- Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *ITL International Journal of Applied Linguistics*, 156, 279-296.
- Van Beuningen, C. G., De Jong, N., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, 62, 1–41.
- VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, 12, 287-301.
- Varnosfadrani, A., & Basturkmen, H. (2009). The effectiveness of implicit and explicit error correction on learners' performance. *System*, 37(1), 82–98.
- Vlugter, P., Knott, A., McDonald, J., & Hall, C. (2009). Dialogue-based CALL: a case study on teaching pronouns. *Computer Assisted Language Learning*, 22(2), 115-131.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

- Vygotsky, L. S. (1987). Thinking and speech. In R.W. Rieber & A.S. Carton (Eds.), *The collected works of L.S. Vygotsky, Volume 1: Problems of general psychology* (pp. 39–285). New York: Plenum Press.
- Wang, P. (2015). Effects of an automated writing evaluation program: student experiences and perceptions. *Electronic Journal of Foreign Language Teaching*, 12(1), 79-100.
- Warschauer, M. (1996). Motivational aspects of using computers for writing and communication. In M. Warschauer (Ed.), *Telecollaboration in foreign language learning: proceeding of the Hawai'i symposium* (pp. 29-46). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Warschauer, M. (2005). Sociocultural perspectives on CALL. In J. Egbert & G. M. Petrie (Eds.) *CALL research perspectives* (pp. 41-51). Mahwah, NJ: Lawrence Earlbaum.
- Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition*, 32, 235-263.
- Yilmaz, Y. (2012). The relative effects of explicit correction and recasts on two target structures via two communication modes. *Language Learning*, 62(4), 1–36.
- Yilmaz, Y. (2013). The relative effectiveness of mixed, explicit and implicit feedback in the acquisition of English articles. *System*, 41, 691-705.
- Zeng, G., & Takatsuka, S. (2009). Text-based peer–peer collaborative dialogue in a computer-mediated learning environment in the EFL context. *System*, 37, 434-446.
- Suzuki, M., (2004). Corrective feedback and learner uptake in adult ESL classrooms. *Columbia University Working Papers in TESOL & Applied Linguistics*, 4, 1-21.

APPENDIX A. VIDEO-STIMULATED RECALL QUESTIONS

1. Here the chatbot highlighted a number of errors in the text, told you that these errors share an error type and asked you if you know this error type. What is your opinion/impression of the way the chatbot is giving you feedback here?
2. After you responded that you know the error type, the chatbot asked you here to select the correct error type from a list of three error types. What is your opinion/impression of being asked to select the correct error type from a list?
3. After you selected the correct error type, the chatbot displayed a sample sentence from the text and asked you to correct it. What is your opinion/impression of this step in the error correction session?
4. After you responded that you do not know the error type, the chatbot started here to explain the error type to you. It first pointed at and told you that that is the first sentence with the simple past tense error. Then, it pointed at and told you that that is the main verb in the sentence and asked you about the tense of the verb. You replied by saying that the verb is in the present tense. Then, the chatbot asked you about the timeframe of the activity being described in the text here. You replied by saying that the activities happened in the past. So, the chatbot asked you about the correct form of the verb in that case. You replied that the main verb should be rewritten to What is your opinion/impression of this way of explaining grammar errors?

APPENDIX B. SURVEY

1. I learn from error corrections how to use grammar rules correctly.

Strongly agree (5) agree (4) Neither (3) disagree (2) Strongly disagree (1)

2. I want to be corrected on all types of grammar errors.

Strongly agree (5) agree (4) Neither (3) disagree (2) Strongly disagree (1)

3. I want to be told about every grammar error in my writing.

Strongly agree (5) agree (4) Neither (3) disagree (2) Strongly disagree (1)

4. When I make an error, I want to be corrected..

- a. Immediately (5)
- b. after done writing the sentence (4)
- c. after done writing the paragraph (3)
- d. after done writing the paper (2)
- e. when I ask for correction (1)

- 5.

- a.** I want the computer to only highlight my grammar errors. For example,

I **eat** lunch at a restaurant **yesterday**.

Strongly agree (5) agree (4) Neither (3) disagree (2) Strongly disagree (1)

- b.** How **clear** is this type of feedback?

Very clear (5) Somewhat clear (4) Neither (3) Somewhat unclear (2)

Very unclear (1)

- c.** How **useful** is this type of feedback?

Very useful (5) Somewhat useful (4) Neither (3) Somewhat unuseful
(2) Very unuseful (1)

6.

a. I want the computer to highlight my grammar errors and tell me the correct forms. For example,

I **eat** lunch at a restaurant **yesterday**.

↓

Correct verb: ate.

Strongly agree (5) agree (4) Neither (3) disagree (2) Strongly disagree (1)

b. How **clear** is this type of feedback?

Very clear (5) Somewhat clear (4) Neither (3) Somewhat unclear (2)

Very unclear (1)

c. How **useful** is this type of feedback?

Very useful (5) Somewhat useful (4) Neither (3) Somewhat unuseful
(2) Very unuseful (1)

7.

a. I want the computer to highlight my grammar errors and tell me their types. For example,

I **eat** lunch at a restaurant **yesterday**.

↓

This verb does not match the timeframe of the sentence. You should use a verb in the

simple past tense.

Strongly agree (5) agree (4) Neither (3) disagree (2) Strongly disagree (1)

b. How **clear** is this type of feedback?

Very clear (5) Somewhat clear (4) Neither (3) Somewhat unclear (2)

Very unclear (1)

c. How **useful** is this type of feedback?

Very useful (5) Somewhat useful (4) Neither (3) Somewhat unuseful

(2) Very unuseful (1)

8.

a. I want the computer to ask me about my grammar errors, ask me to fix them and give me hints when I need help. For example,

I **eat** lunch at a restaurant **yesterday**.

↓

Computer: You have a grammar error. Do you know what type is it?

You: No.

Computer: Is it a simple past, a present perfect, or a present progressive error?

You:...

Strongly agree (5) agree (4) Neither (3) disagree (2) Strongly disagree (1)

b. How **clear** is this type of feedback?

Very clear (5) Somewhat clear (4) Neither (3) Somewhat unclear (2)

Very unclear (1)

c. How **useful** is this type of feedback?

Very useful (5) Somewhat useful (4) Neither (3) Somewhat unuseful
(2) Very unuseful (1)

9. How often do you use automated writing evaluation software that show and correct your errors?

Always (5) Often (4) Sometimes (3) Rarely (2)
Never (1)

10. I have used the following automated writing evaluation software:

ETS Criterion MY Access! add choice...

a. Please rate your overall experience with ETS Criterion

Very positive (5) Positive (4) Neither (3) Negative (2)
Very Negative (1)

b. Please rate your overall experience with MY Access!

Very positive (5) Positive (4) Neither (3) Negative (2)
Very Negative (1)

APPENDIX C. INTERVIEW QUESTIONS

1. In item 1 on the survey, you have indicated that you learn from error correction how to use grammar rules.

Can you tell me why you think error correction helps you learn?

2. In item 2 and 3, you have indicated that you want to be corrected for all grammar error types and for every error.

Why do you want to be corrected for every type and for every error?

How would this affect your focus on expressing your ideas?

3. In item 4, you have indicated that you want to be corrected immediately/ after finishing your paper/ etc.

Why do you prefer to be corrected immediately/ after finishing your paper/ etc?

4. According to items 5, 6, 7 and 8, you prefer correction strategy 5 / 6 / 7 / 8 and you think that strategy 5 / 6 / 7 / 8 is more clear and useful.

Why do you prefer this strategy over others? Why do you think strategy is more clear and useful than others?

5. What is your overall opinion of *wTutor*? Do you see any weakness or advantage to the type of error correction provided by *wTutor*? Is there anything that you would like to see changed, removed or added to *wTutor*?

APPENDIX D. REPOSITORIES, PUBLICATIONS, AND KEYWORDS USED IN LITERATURE SURVEY

1. List of academic publication repositories searched for primary CF studies

- DART-Europe E-theses Portal
- ERIC Institute of Education Sciences
- ProQuest - Linguistics and Language Behavior Abstracts (LLBA)
- Taylor & Francis Group
- Wiley Online Library

2. List of peer-reviewed journals searched for primary CF studies

- Applied Linguistics
- Computer Assisted Language Learning
- Foreign Language Annals
- Innovation in Language Learning and Teaching
- International Journal of Applied Linguistics
- International Journal of Pedagogies and Learning
- International Review of Applied Linguistics
- Journal of Second Language Writing
- Language Awareness
- Language Learning
- Language Learning and Technology
- Language Teaching
- Research Language, Culture and Curriculum Studies in Second Language Acquisition
- System

- TESOL Quarterly
- The Canadian Modern Language Review
- The Journal of Language Teaching and Learning
- The Language Learning Journal
- The Modern Language Journal
- World Journal of Education

3. Search keys used in searching for primary CF studies

- corrective feedback
- explicit feedback
- focus on form
- focus on forms
- implicit feedback
- incidental language learning
- metalinguistic feedback
- prompts
- recasts

APPENDIX E. FORTY-FOUR CF STUDIES QUALIFIED FOR A META-ANALYSIS

1. Twenty-Five Oral CF Studies Qualified for a Meta-Analysis

Oral CF Studies	Recast	Prompt	Metalinguistic	Clarification R.	Repetition
Ammar & Spada, 2006	✓	✓			
Ellis et al., 2006	✓		✓		
Ellis, 2007	✓		✓		
Faqeih, 2012 (FL)	✓		✓		
Faqeih, 2012 (SL)	✓		✓		
Kartchava & Ammar, 2014	✓	✓			
Leeman, 2003	✓				✓
Li, 2013	✓				
Long et al., 1998	✓				
Loewen & Nabei, 2007	✓		✓	✓	
Lyster, 2004	✓	✓			
Mackey & Philp, 1998	✓				
Monteiro, 2014	✓		✓		
Profozic, 2013	✓			✓	
Rahimi & Zhang, 2013	✓				
Rassaei & Moinzadeh, 2011	✓			✓	
Rassaei et al., 2012	✓		✓		
Rassaei, 2013	✓				
Rassaei, 2014	✓				
Rassaei, 2015a	✓		✓		
Sheen, 2007a	✓		✓		

Sheen, 2008	✓		
Takashima & Ellis, 1999			✓
Yang & Lyster, 2010	✓	✓	
Yilmaz, 2013	✓		

2. Nineteen Written CF Studies Qualified for a Meta-Analysis

Oral CF Studies	Direct Correction	Metalinguistic	Indirect Correction
Bitchener & Knoch, 2008	✓	✓	
Bitchener & Knoch, 2009	✓	✓	
Bitchener & Knoch, 2010		✓	✓
Bitchener, 2008	✓	✓	
Ellis et al., 2008	✓		
Farrokhi & Sattarpour, 2012	✓		
Jiang & Xiao, 2014	✓	✓	
Mubarak, 2013	✓	✓	
Parreno, 2013	✓	✓	✓
Sheen et al., 2009	✓		
Sheen, 2007b	✓	✓	
Sheen, 2010b	✓	✓	
Shintani & Ellis, 2013	✓	✓	
Stefanous & Révész, 2015	✓	✓	
Van Beuningen et al., 2008	✓		✓
Van Beuningen et al., 2012	✓		✓
Sheen, 2010b	✓	✓	

Bitchener & Knoch, 2010

✓

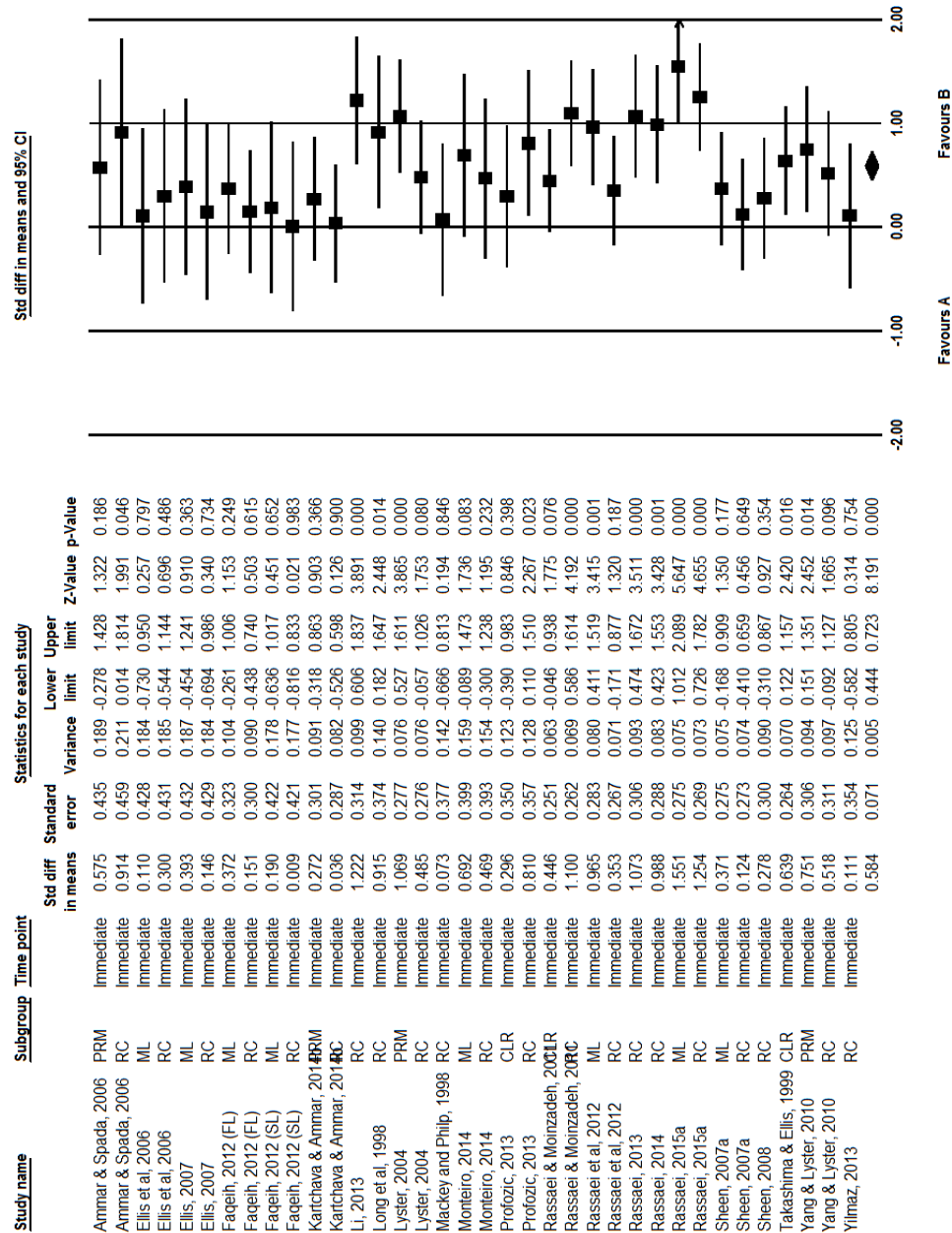
✓

Frear & Chiu, 2015

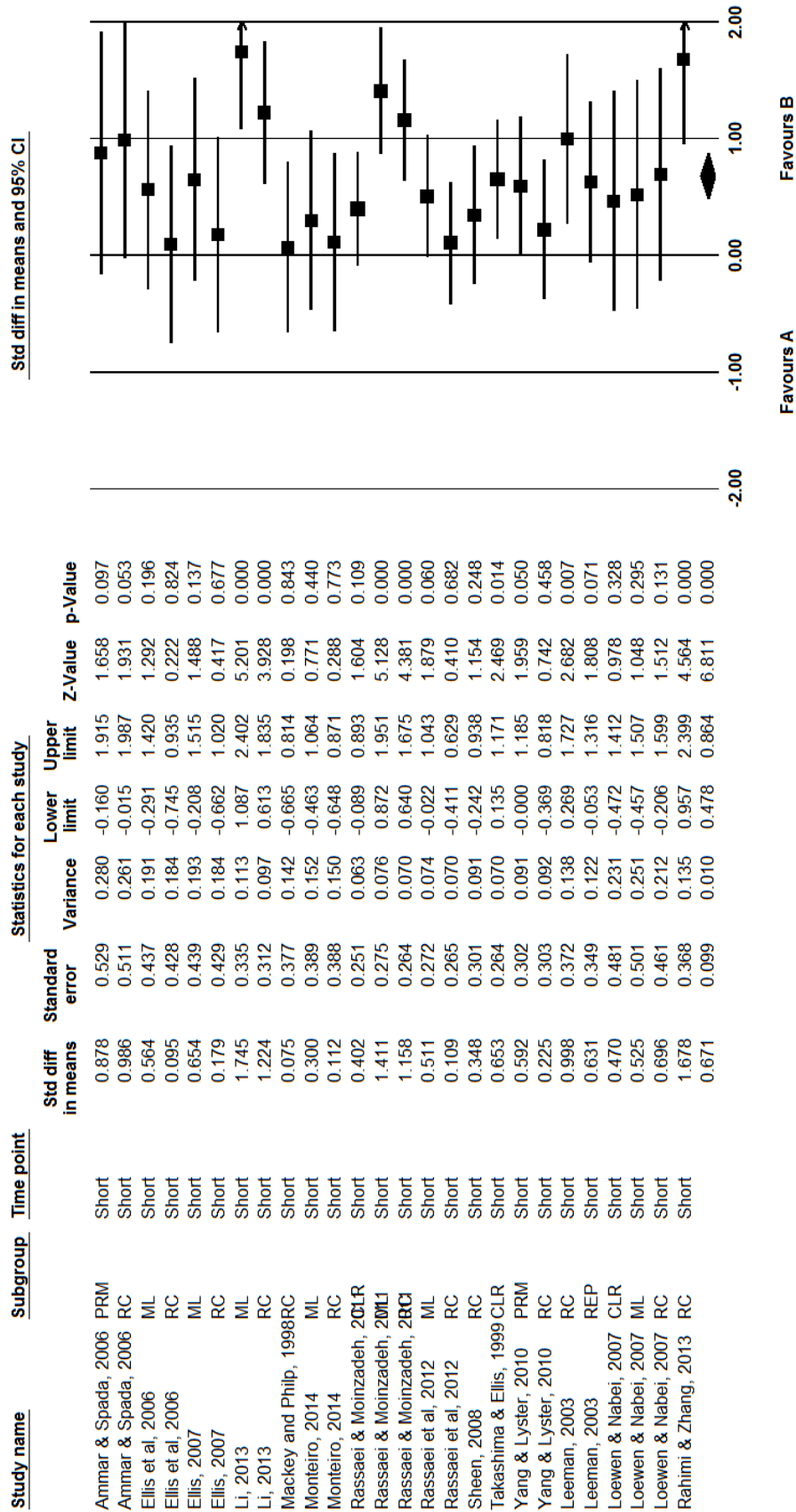
✓

APPENDIX F. META-ANALYSIS RESULTS

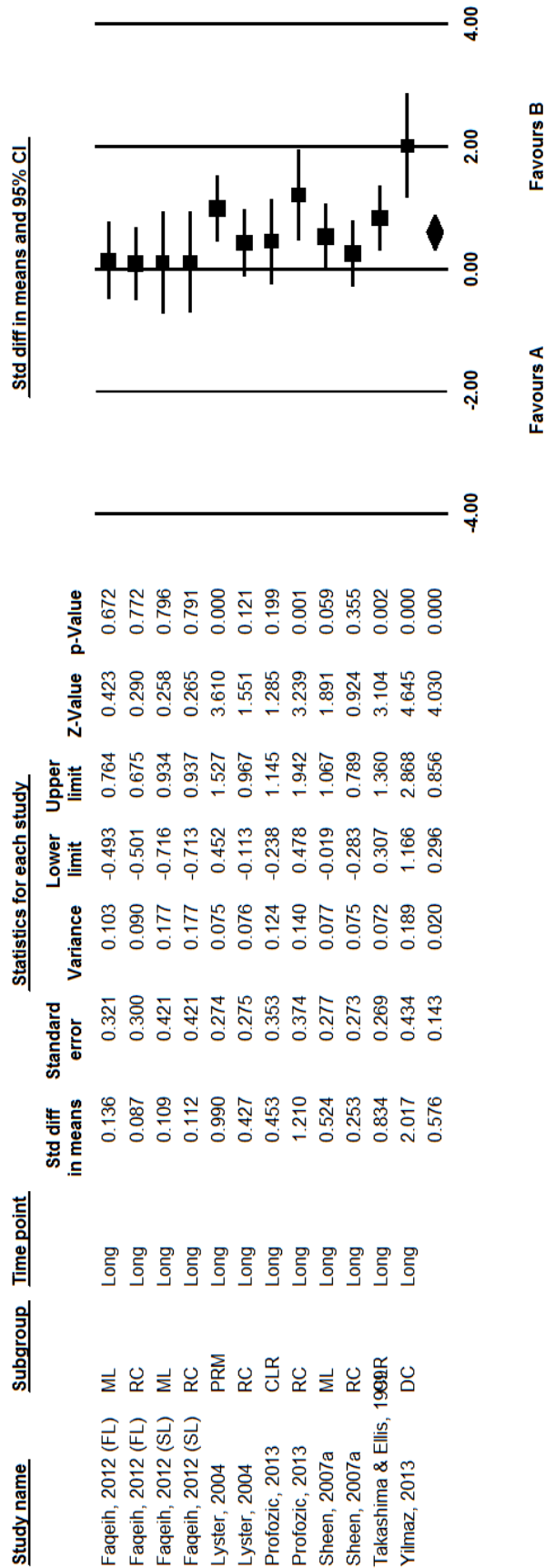
Immediate Effect of Oral Corrective Feedback



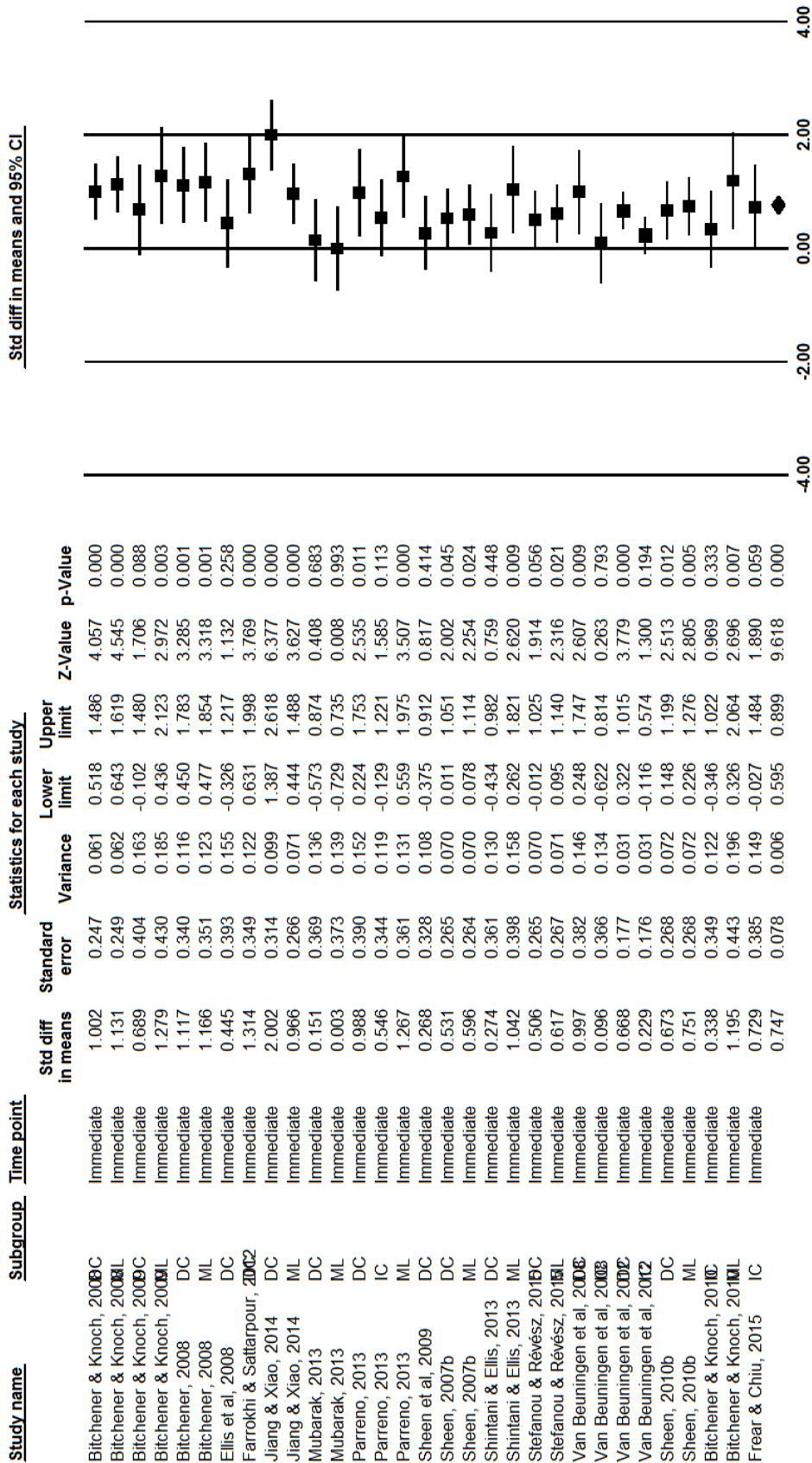
Short Effect of Oral Corrective Feedback



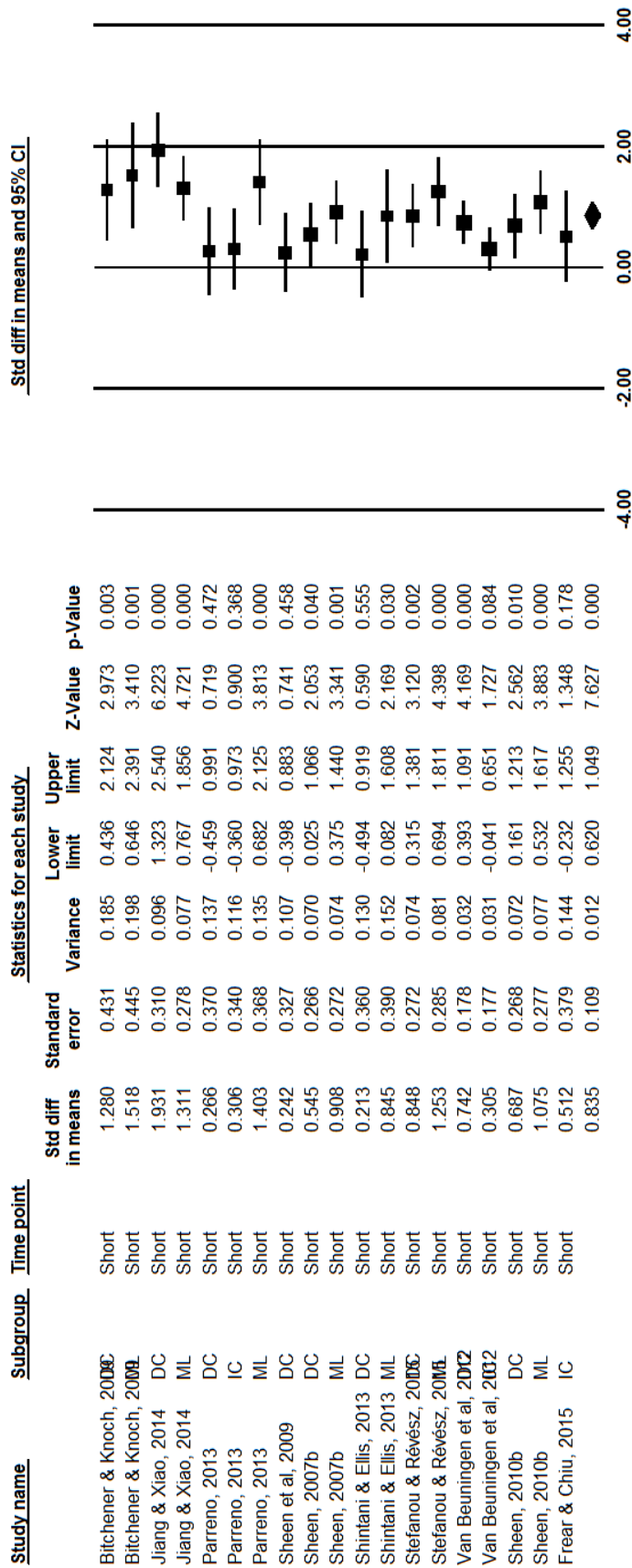
Long Effect of Oral Corrective Feedback



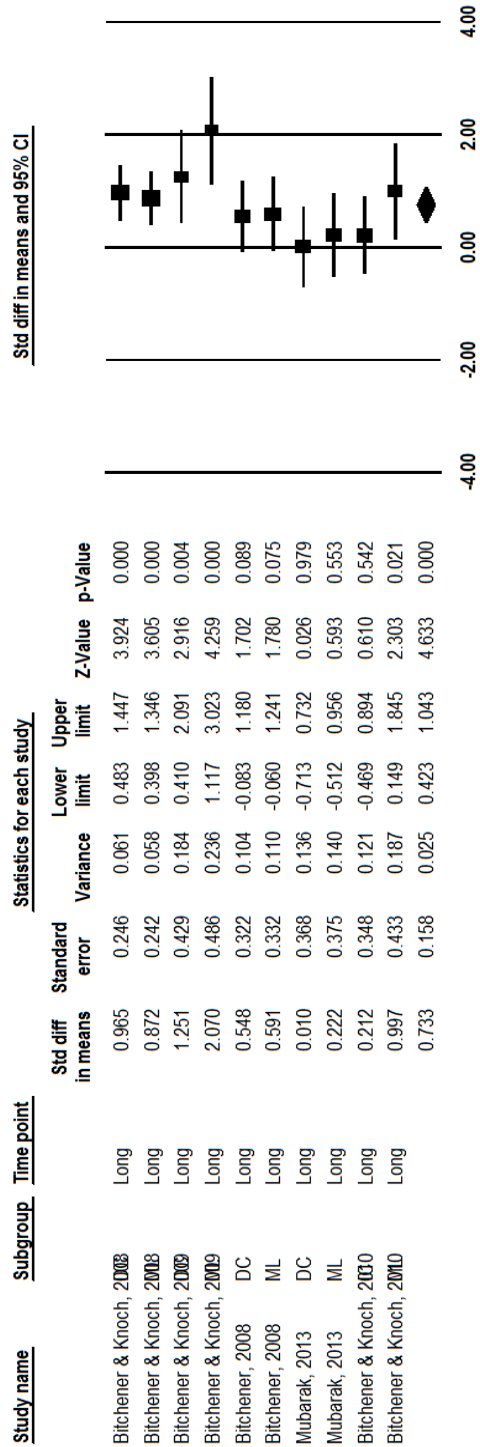
Immediate Effect of Written Corrective Feedback



Short Effect of Written Corrective Feedback



Long Effect of Written Corrective Feedback



APPENDIX G. PERCEPTIONS AND PREFERENCES OF CF

Item	Explanation	Category	#
CF usefulness	CF is useful because it lets you know the error and its type.	Noticing Learning	10
	CF boosts writers' confidence in the quality of their writing.	Confidence	1
	CF helps writers avoid making similar errors in the future.	Avoid future errors	5
	CF improves writing quality.	Writing quality	1
	CF reminds writers of correct grammar.	Remember	3
	CF promotes self-reliance and independence.	Independence	2
	Writers may over-rely on automated CF for finding mistakes.	Independence	1
	CF pushes writers to think about errors and how to correct them.	Learning	1
	CF is not helpful because of inaccurate error detection.	Accuracy of error detection	3
CF frequency	CF may not lead to learning when rushing to finish paper and have no time to think about errors.	Time	1
	A lot of grammar-related CF can interrupt focus on writing.	Interruption of idea flows	4
	Serious grammar error types should be prioritized over less serious types.	Prioritize CF based on error type	1
	Less frequent CF is preferred because automated error detection is not very accurate.	Accuracy of automated error detection	2
	Comprehensive CF is need in order to identify and correct all mistakes.	Accuracy	6
	Comprehensive CF may lead learners to over rely on software for	Self-reliance	1

	finding errors and not learn from correcting their own errors.		
Timing	Immediate CF helps writers avoid repeating the same kinds of errors in the rest of their papers.	Immediate CF	1
	Immediate CF is better for learning/remembering.	Immediate CF	2
	Immediate CF saves time by preventing writers from repeating same mistakes.	Immediate CF	1
	Immediate CF delivered as identification does not interrupts expression of ideas especially with short prose.	Immediate CF	2
Immediate CF	Immediate CF is easier to see and understand.	ease	1
	Immediate CF may interrupts focus on meaning.	Interruption of ideas	3
	Helps in learning from errors because of short time between error and correction.	Learning	2
	Improve immediate subsequent writing by preventing repeated errors.	Accuracy	2
	Immediate CF is desired when can be ignored and addressed later.	Delayable response	2
	Helps focus more on writing.	Focus	1
CF per sentence	CF per sentence helps writers avoid repeating the same kinds of errors in the rest of their papers.	Immediately improve subsequent writing quality	2
	CF per sentence does not overwhelm writers with CF as the case with CF per paragraph or at the end of writing.	Non-overwhelming amount of CF	1
	CF per sentence helps in expressing ideas before receiving grammar	Interruption of ideas	1

related CF.

	CF per sentence sometimes interrupts expression of ideas when an idea is expressed over more than a single sentence.	Interruption of ideas	1
	CF per sentences does not interrupt expression of ideas for some learners who are still struggling with grammar and mechanics.	Interruption of ideas	1
	A sentence is the minimum amount of text that can be analyzed accurately by a computer.	Accuracy of automated CF	1
	CF per sentence acts as a quality checker on sentence by sentence basis before forgetting the sentence.	Writing accuracy	1
CF per paragraph	A paragraph expresses a complete idea and receiving CF after completing writing a paragraph does not interrupt focus on meaning.	Interruption of ideas	3
	Receiving CF per paragraph could be overwhelming if many mistakes exist.	Overwhelming	1
	A paragraph is a reasonable amount of text to review and correct.	Overwhelming	2
	It saves more time to receive CF per paragraph then per sentence.	Saves time	1
CF after finish writing	Receiving CF after done writing a paragraph helps writers focus on meaning and organization.	Interruption of ideas	4
On-Demand CF	Allowing writers to trigger CF helps them prioritize other aspects of writing.	Interruption of ideas	5
Identification	Identification alone is not very useful because it does not tell writers why the highlighted text is incorrect.	Usefulness	1
	Identification brings writers'	Attention	4

	attention to errors.		
	Identification is not enough to make the right corrections.	Clarity	8
	Identification is most useful for advanced ESL writers.	Proficiency level	1
Direct Correction	Direct correction saves time by not requiring writers to find and type corrections themselves.	Save time	7
	Direct correction does not distract writers too much from their other writing goals.	Less distraction	1
	Direct correction does not help writers learn how to correct their mistakes.	Learning	6
	Direct correction is better than identification alone because it leads to learning.	Learning	1
	Direct correction is useful when the correct answer is unknown.	When answer is unknown	1
	Direct correction is easier, simpler and less prone to confusion than other CF strategies. Error and correction can be clearly seen next to each other. It leaves no room for guessing.	Ease	5
Metalinguistic CF	Metalinguistic CF is useful for learning about errors and how to correct them.	Learning	13
	Metalinguistic CF is not desired when a writer wants to finish a paper as soon as possible.	Waste time	5
	Metalinguistic CF allows writers to decide if automated error detection is accurate or not.	Accuracy of automated error detection	1
	Metalinguistic CF pushes writers to think about their errors.	Promote thinking	1

	Metalinguistic CF may be more helpful for beginner ESL writers because it teaches them how to correct their errors. Advanced writers do not need information about the error.	Better suited for beginners	4
	Metalinguistic CF makes correcting error harder for writers. Unfamiliar meta-language could make correction harder.	Harder	5
	Metalinguistic CF leads writers to avoid similar errors in subsequent writing.	Improve quality	1
	Metalinguistic CF should be contingent on error type and should not be repeated for every instance of the same error type. Metalinguistic CF should be simpler for simple basic errors.	Dependent on error type	2
Graduated CF	Graduated CF gives writers chance and push them to think about and remember the correct grammar rule.	Remember	8
	Graduated CF is good for learning. It helps learn about grammar rules. Questions are helpful for learning. Interactivity and guidance is helpful. Delaying giving the correct answer is helpful.	Learning	15
	Graduated CF does not lead to learning because it is in the form of questions. Questions do not help in learning if writers do not know the answers.	Learning	1
	Graduated CF may not accept a different but valid answer.	Flexibility	1
	Graduated CF promote learner autonomy.	Learner autonomy	1
	Graduated CF is too complicated.	Complexity	1

	Graduated CF involves too much effort.	Difficulty	1
	Graduated CF is not suitable for absolute beginners who do not have enough grammar knowledge.	Appropriate proficiency level	1
	Graduated CF is suitable for beginners as it acts like a tutor and provide guidance. Advanced writers just need quick reminders.	Appropriate proficiency level	3
	Desirability depends on whether the writer wants to learn or just have errors corrected.	Desirability	3
	Language of graduated CF should be suitable for the proficiency level of writers.	Language	1
	Graduated CF is interesting because it resembles interaction with human tutors.	Human-like interaction	1
	Graduated CF requires too much time and effort especially if writers are only focusing on correcting their errors.	Time	14
Multiple choices as CF strategy	Giving CF as multiple choices pushes learners to think about their errors.	Prompt thinking	1
Tooltip-based CF	CF should be presented when writers mouseover highlighted errors.	On-demand CF	1
Metalinguage	Metalinguage used in explanation and graduated CF could be unfamiliar for some writers.	Familiarity with meta-language	6
	Metalinguage should be explained or translated to native language of writer.	Simplify meta-language	3

APPENDIX H. PERCEPTIONS OF GRADUATED CF AND WTUTOR

Interaction	Belief/Preference	Category	#
Graphical interface / User experience	Text font should be larger even if that led to smaller avatar.	Text font	4
	Text-to-speech playback of chatbot dialog may disrupt some learners who prefer visual input.	Attention	1
Identifying errors of the same type	This saves writers times because writers do not need to go through text again to find the same error type.	Time	1
	Identification does not push writers to find their own errors and focus on writing more accurate text.	Learn autonomy	2
	Telling writers the number of errors instead of identification make them focus more on their text.		
	Identification focuses writers' attention on errors.	Attention	1
Question about error type	Receiving correction after incorrect responses causes writers to remember error better.	Learning	1
	"if the computer give me the answer.. easy come easy go"		
	Questioning pushes writers to recall what they may already know about the error and learn from their mistakes.	Learning	5
	This pushes writers to actively think about the error and its type.	Learning	3
	"...when it said wrong I tried to get my brain get more focus and that's when I realize Ooooh this something should probably be with a tense. "		
	Questions are better than auto-correction.	Graduated vs. direct correction	2

Prompt to select answer from list	List of choices acts like refresher and narrows down possible error types.	Ease, Remember	3
	Selection helps writers remember the name of target grammar.		
	Exclusive use of selection list instead of typing answers makes CF faster and easier.	Selection vs. typing	1
	Selecting correct answer requires basic knowledge about target grammar.	Proficiency level	1
	Prompting writers to select the correct error type from a list when they do not provide the correct error type gives writers several opportunities to correct themselves.	Multiple opportunities to self-correct	1
Prompt to correct example sentence	Getting the opportunity to correct an example sentence and making a mistake causes writers to remember the correct answer.	Learning	1
	This helps writers test their knowledge about target language. It helps writers practice what they have learned while providing immediate feedback.	Knowledge testing Practice	3
	This pushes writers to consciously reflect and correct their errors.	Conscious reflection Thinking	2
	This makes writers think and try to correct their error.		
	Example sentence needs more context or instruction to make the purpose of this step clear for writers.	Graphical interface / User experience	2
	Correcting example sentence inside the original paragraph is easier and faster.	Graphical interface / User experience	2
	For low proficiency writers, a list of choices is better than typing a	Graphical interface / User experience	2

correction.

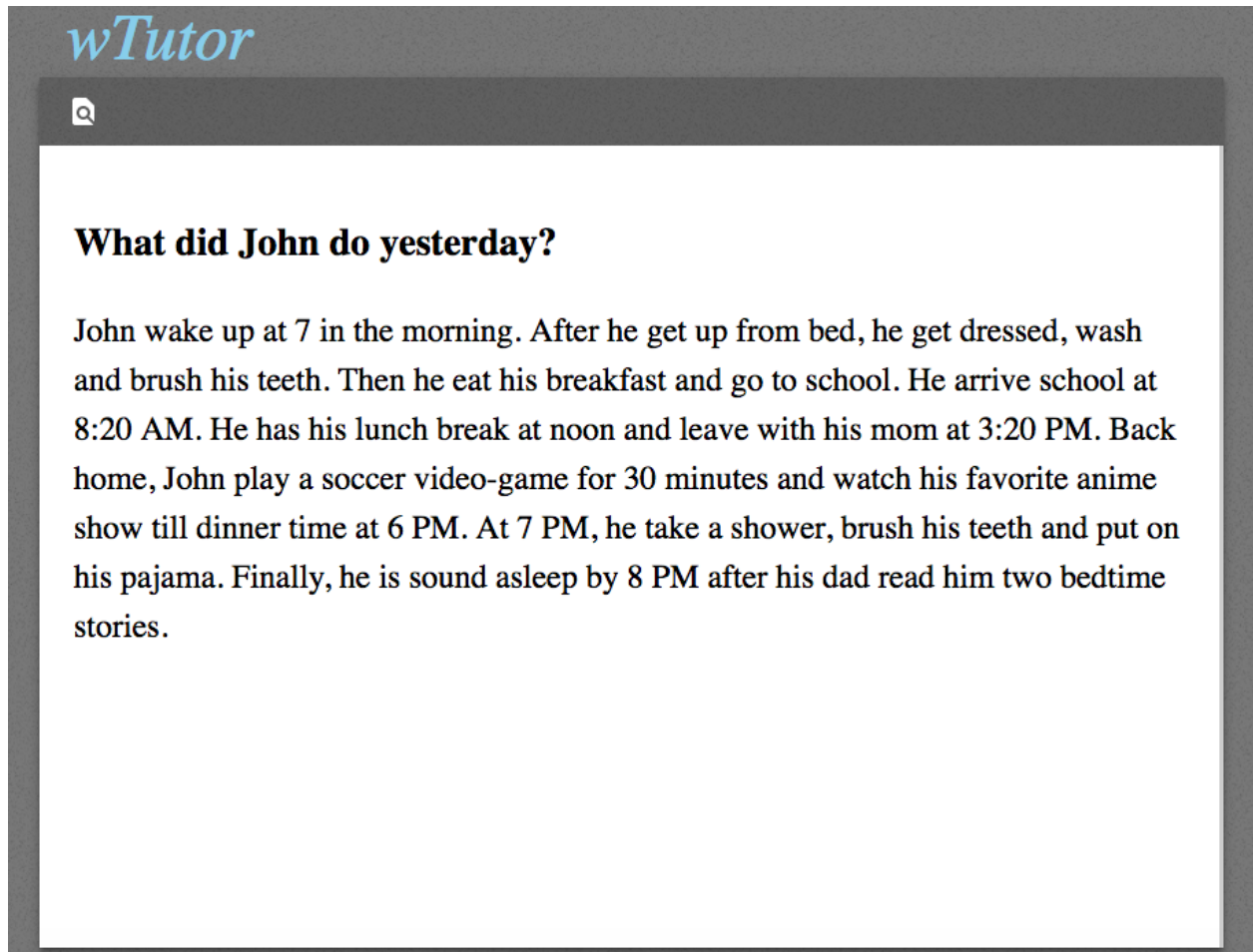
I am concerned about the tool's ability to accurately evaluate my response. Therefore, a list of choices would be more helpful.

Multiple correction attempts	Allowing writers multiple opportunities to respond correctly helps information "stick in mind"	Learning	2
	Multiple attempts ease writers concern about getting stuck in dialog with the computer tutor.	Concern about being stuck	2
Explanation of grammar rule with questions	The steps used in explaining the target grammar rule helps reminds writers of the correct words to use.	Learning	2
	It pushes writers to think about why the highlighted text is incorrect. "The procedure helped yes, narrowed my thinking"	Learning	4
	"I have confidence even if I do wrong, it's going to help explaining, until I get it right".	Confidence	1
	Beginners can learn from explanation.	Learning	1
	"...explanation helps you to correct it by yourself."	Learner autonomy	2
	Receiving correction after incorrect attempts focuses writer's attention on correction.	Attention	1
	Explanation should be more specific and detailed. For example, explanation should have included the distinction between regular and irregular past verb forms.	Detailed explanation	1
Graduated CF suitability for advanced language features	Graduated CF may not be suitable for correcting and explaining advanced language features.	Target language level	1

Interaction with <i>wTutor</i>	Interaction with the tool is easy and friendly.	Ease of interaction	2
	Graduated CF interaction is harder than direct correction because it requires more thinking and concentration.	Ease of interaction	1
	Interaction should be simple and faster for information to be remembered. For example, after one incorrect try, writers should get the correct answer immediately.	Complexity	1
	Successful and helpful interaction with the tool requires certain level of ESL proficiency.	Proficiency level	2
	This interaction is good for learning. This interaction gives writers “opportunity to practice and learn”.	Learning	8
	The tool can be used by writers to learn by themselves.	Learner autonomy	1
	Interaction should be altered based on how long a user has used the tool and how many corrections a user has received on the same error type. For example, experienced user may only need a quick reminder of the types of errors in their text.	Dynamic interaction	1
	This interaction requires a lot of time for correction.	Time	7
	Writers may not have enough time for this interaction when they are writing for a content course.		
Tool language level	This interaction may take too much time for correcting long text.		
	This interaction can waste time if repeated.		
	The tool should accommodate the proficiency level of target learners by	Proficiency level	3


	using simple language with low proficiency learners.		
	Metalanguage could be difficult and unfamiliar to some writers.	Metalanguage	13
	Metalanguage should be explained using tooltip with definition or translation.	Metalanguage	5
L2 proficiency level	Graduated CF may be more useful for beginners especially because it offers opportunity to practice.	Proficiency level	6

APPENDIX I. WTUTOR SCREENSHOTS

1. Initial state of *wTutor* editor


2. *wTutor* reporting error

wTutor




What did John do yesterday?

John **wake** up at 7 in the morning. After he **get** up from bed, he **get** dressed, **wash** and **brush** his teeth. Then he **eat** his breakfast and **go** to school. He **arrive** school at 8:20 AM. He has his lunch break at noon and leave with his mom at 3:20 PM. Back home, John **play** a soccer video-game for 30 minutes and **watch** his favorite anime show till dinner time at 6 PM. At 7 PM, he **take** a shower, **brush** his teeth and put on his pajama. Finally, he **is** sound asleep by 8 PM after his dad read him two bedtime stories.





wTutor: The 15 highlighted words have the same type of error. Can you tell me the type of these errors?

chat...

 verb

3. *wTutor* explaining error type





What did John do yesterday?

John **wake** up at 7 in the morning. After he **get** up from bed, he **get** dressed, **wash** and **brush** his teeth. Then he **eat** his breakfast and **go** to school. He **arrive** school at 8:20 AM. He has his lunch break at noon and leave with his mom at 3:20 PM. Back home, John **play** a soccer video-game for 30 minutes and **watch** his favorite anime show till dinner time at 6 PM. At 7 PM, he **take** a shower, **brush** his teeth and put on his pajama. Finally, he **is** sound asleep by 8 PM after his dad read him two bedtime stories.

wTutor: The 15 highlighted words have the same type of error. Can you tell me the type of these errors?

You: no

wTutor: Alright. I will explain this grammar error to you.

wTutor: This is the first example of simple past error.

wTutor: "wake" is the main verb in the sentence.

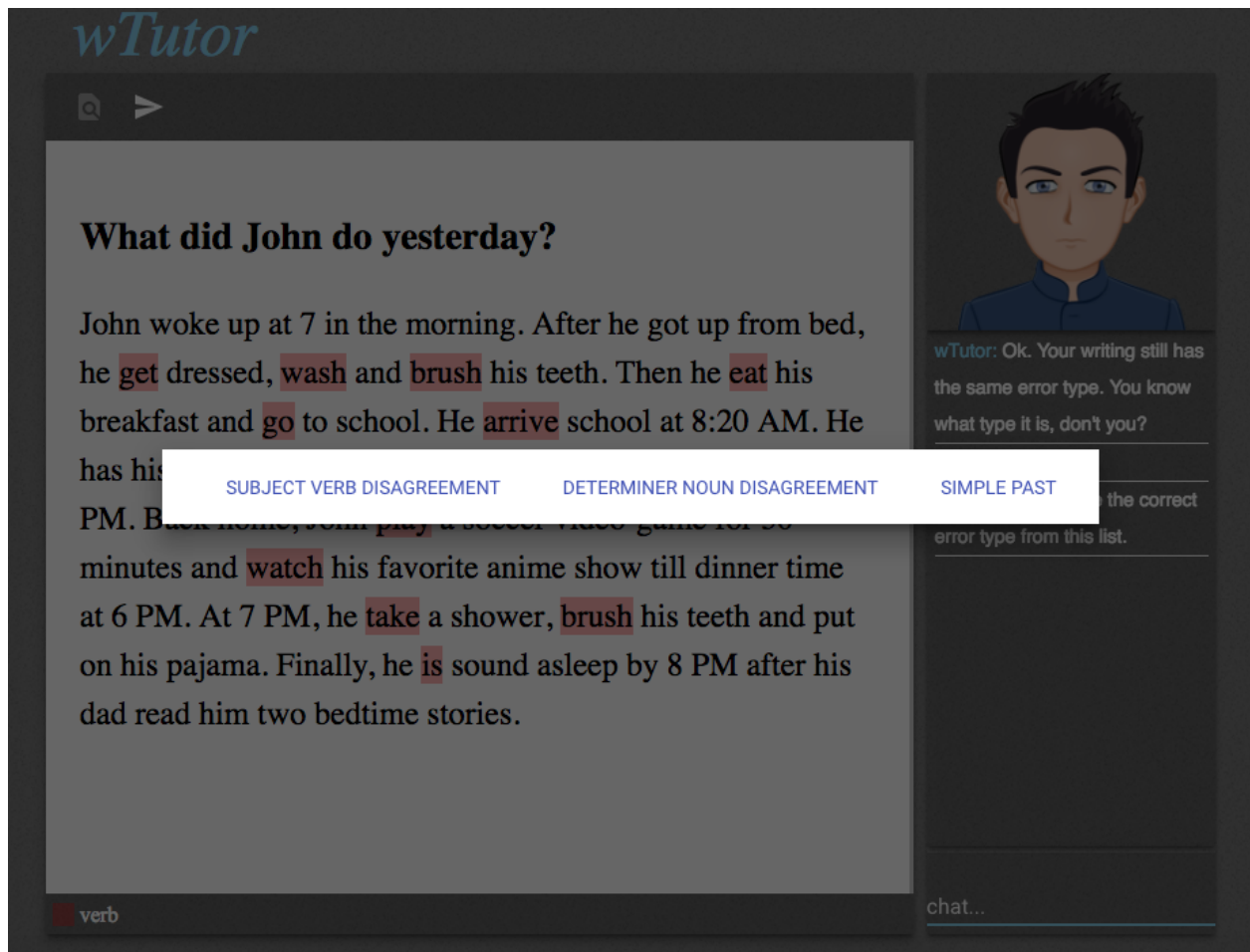
wTutor: Tell me the tense of this verb.

verb

chat...

4. *wTutor* evaluating knowledge about error type

4.1. *wTutor* prompting user to select error type



The screenshot shows the *wTutor* interface. On the left, a text box contains a paragraph about John's day. Several words are highlighted in red boxes: **get**, **wash**, **brush**, **eat**, **go**, **arrive**, **watch**, **take**, and **is**. A white menu box is overlaid on the text, containing three options: **SUBJECT VERB DISAGREEMENT**, **DETERMINER NOUN DISAGREEMENT**, and **SIMPLE PAST**. On the right, there is a chat window with a cartoon avatar of a man. The chat text reads: "wTutor: Ok. Your writing still has the same error type. You know what type it is, don't you?". Below the chat text is a text input field with the placeholder "chat...".

What did John do yesterday?

John woke up at 7 in the morning. After he got up from bed, he **get** dressed, **wash** and **brush** his teeth. Then he **eat** his breakfast and **go** to school. He **arrive** school at 8:20 AM. He has his PM. Back home, John **play** a soccer game for 30 minutes and **watch** his favorite anime show till dinner time at 6 PM. At 7 PM, he **take** a shower, **brush** his teeth and put on his pajama. Finally, he **is** sound asleep by 8 PM after his dad read him two bedtime stories.


SUBJECT VERB DISAGREEMENT DETERMINER NOUN DISAGREEMENT SIMPLE PAST



wTutor: Ok. Your writing still has the same error type. You know what type it is, don't you?

the correct error type from this list.

chat...


4.2. wTutor prompting user to correct sample sentence



What did John do yesterday?

John woke up at 7 in the morning. After he got up from bed, he **get** dressed, **wash** and **brush** his teeth. Then he **eat** his breakfast and **go** to school. He has his lunch break at noon. At 1 PM. Back home, John plays for 30 minutes and **watch** his favorite TV show. At 6 PM. At 7 PM, he **take** a shower, **brush** his teeth and put on his pajama. Finally, he **is** sound asleep by 8 PM after his dad read him two bedtime stories.



wTutor: Ok. Your writing still has the same error type. You know what type it is, don't you?

You: yes

wTutor: Ok. Choose the correct error type from this list.

wTutor: Correct!

wTutor: Now correct this sentence.


chat...



verb

he get dressed

5. *wTutor* ending dialog


5.1. *wTutor* prompting user to correct similar errors




What did John do yesterday?

John woke up at 7 in the morning. After he **got** up from bed, he **get** dressed, **wash** and **brush** his teeth. Then he **eat** his breakfast and **go** to school. He **arrive** school at 8:20 AM. He has his lunch break at noon and leave with his mom at 3:20 PM. Back home, John **play** a soccer video-game for 30 minutes and **watch** his favorite anime show till dinner time at 6 PM. At 7 PM, he **take** a shower, **brush** his teeth and put on his pajama. Finally, he **is** sound asleep by 8 PM after his dad read him two bedtime stories.

 verb



wTutor: I found a common type of error in your writing. Do you know the error type in the highlighted words?

You: simple past

wTutor: Correct!

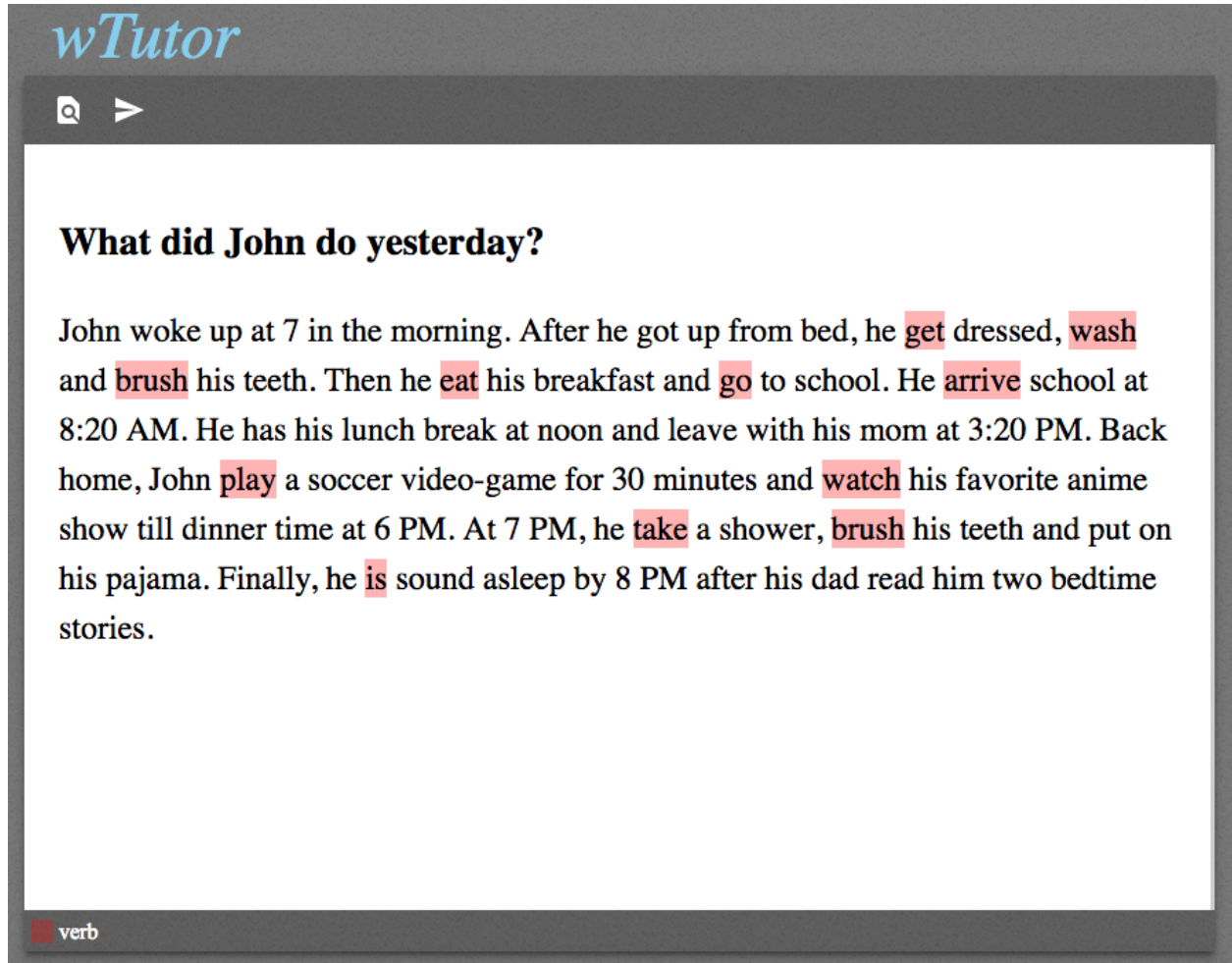
wTutor: Now correct this sentence.

wTutor: Excellent.

wTutor: Now, correct similar errors in your writing.

chat...

5.2. State of editor after wTutor ends dialog



The screenshot shows the wTutor editor interface. At the top left, the logo "wTutor" is displayed in a stylized blue font. Below the logo is a dark gray toolbar containing a magnifying glass icon and a right-pointing arrow icon. The main area is a white text editor with the following text:

What did John do yesterday?

John woke up at 7 in the morning. After he got up from bed, he **get** dressed, **wash** and **brush** his teeth. Then he **eat** his breakfast and **go** to school. He **arrive** school at 8:20 AM. He has his lunch break at noon and leave with his mom at 3:20 PM. Back home, John **play** a soccer video-game for 30 minutes and **watch** his favorite anime show till dinner time at 6 PM. At 7 PM, he **take** a shower, **brush** his teeth and put on his pajama. Finally, he **is** sound asleep by 8 PM after his dad read him two bedtime stories.

At the bottom left of the editor, there is a small red square icon followed by the text "verb".

APPENDIX J. PLOT FOR EXPLORATORY FACTOR ANALYSIS

	cf_usefulness	freq_type	freq_error	timing	ident_desirability	ident_clarity	ident_usefulness	correct_desirability	correct_clarity	correct_usefulness	meta_desirability	meta_clarity	meta_usefulness	grad_clarity	grad_usefulness	freq_use
cf_usefulness	1	-0.15	-0.11	0.21	-0.03	0.13	-0.11	0.03	0.02	0	-0.03	-0.17	-0.14	-0.09	0.1	-0.08
freq_type	-0.15	1	0.67	0.09	-0.29	0.1	-0.18	0.44	-0.09	0.4	0.17	0.3	0.15	0.26	0.45	0.2
freq_error	-0.11	0.67	1	0.05	-0.21	0.23	-0.09	0.38	-0.22	0.54	-0.2	0.08	-0.15	0.23	0.09	0.05
timing	0.21	0.09	0.05	1	-0.04	0.1	0.11	0.13	-0.17	0.06	-0.42	-0.24	-0.26	-0.07	-0.11	-0.03
ident_desirability	-0.03	-0.29	-0.21	-0.04	1	0.51	0.62	-0.02	0.22	0.1	-0.13	0.02	0.1	-0.01	-0.21	-0.33
ident_clarity	0.13	0.1	0.23	0.1	0.51	1	0.4	0.22	0.29	0.66	-0.26	0.02	-0.02	0.25	0.04	-0.26
ident_usefulness	-0.11	-0.18	-0.09	0.11	0.62	0.4	1	-0.13	0.11	0.23	-0.18	-0.08	0.16	-0.03	-0.05	-0.16
correct_desirability	0.03	0.44	0.38	0.13	-0.02	0.22	-0.13	1	0.16	0.65	0.12	0.34	0.1	-0.13	0.47	0.37
correct_clarity	0.02	-0.09	-0.22	-0.17	0.22	0.29	0.11	0.16	1	0.22	0.34	0.53	0.11	0.41	0.31	-0.02
correct_usefulness	0	0.4	0.54	0.06	0.1	0.66	0.23	0.65	0.22	1	-0.11	-0.11	0.01	0.45	0.02	0.02
meta_desirability	-0.03	0.17	-0.2	-0.42	-0.13	-0.26	-0.18	0.12	0.34	-0.11	1	0.56	0.7	0.56	0.23	0.23
meta_clarity	-0.17	0.3	0.08	-0.24	0.02	0.02	-0.08	0.32	0.53	0.11	0.56	1	0.68	0.37	0.34	0.42
meta_usefulness	-0.14	0.15	-0.15	-0.26	0.1	-0.02	0.16	0.1	0.34	0.01	0.7	0.68	1	0.15	0.52	0.16
grad_clarity	-0.09	0.26	0.23	-0.07	-0.01	0.25	-0.03	-0.13	0.41	0.08	0.18	0.37	0.15	1	0.16	-0.2
grad_usefulness	0.1	0.45	0.09	-0.11	-0.21	0.04	-0.05	0.47	0.31	0.45	0.56	0.34	0.16	0.16	1	0.16
freq_use	-0.08	0.2	0.05	-0.03	-0.33	-0.26	-0.16	0.37	-0.02	0.02	0.23	0.42	0.16	-0.2	0.16	1

APPENDIX K. R SCRIPT OF DESCRIPTIVE STATISTICS AND EXPLORATORY FACTOR ANALYSIS

```
# load survey data #####

surveys <- read.table("~/surveys.txt",
                      header=TRUE,
                      sep=";",
                      row.names="name")[1:17]

#####

# calculate descriptive statistics #####

library(psych)

describe(surveys)

# calculate frequency count and percentage of responses for each item #####

for (item in names(surveys)) {
  count <- table(surveys[item])
  percent <- prop.table(count)
  countPercent <- rbind(count, percent)
  print(item)
  print(countPercent)
}

#####

# Find the Kaiser, Meyer, Olkin (KMO) Measure of Sampling Adequacy (MSA) ###
# (Cerny & Kaiser, 1977; Kaiser, 1974)

KMO(surveys) # Overall MSA = 0.44 but minimum accepted value is 0.50

# Does removing one item raise MSA above 0.49?

for(i in names(surveys)){
  da <- surveys
  da[i] <- NULL
  re <- KMO(da);
  print(i)
  print(re$MSA)
```

```

}

# Yes. Removing item 8a (desirability for graduated CF) raises MSA to 0.51
surveys$grad_desirability <- NULL # remove item 8a

KMO(surveys) # Overall MSA = 0.51

#####

# Bartlett's sphericity test #####

# Tests the hypothesis that correlations between variables are greater than
# would be expected by chance

cortest.bartlett(surveys, n=30) # chisq = 238.2672, p.value = 7.856561e-10, df = 120

#####

# Determine Number of Factors to Extract #####

library(nFactors)

ev <- eigen(cor(surveys)) # get eigenvalues

ap <- parallel(subject = nrow(surveys), var=ncol(surveys), rep=100, cent=.05)

nS <- nScree(x = ev$values, aparallel = ap$eigen$qevpea)

plotnScree(nS, xlab = "Factors", main = "")

# 3 factors can be extracted from surveys

#####

# Exploratory Factor analysis #####

fit <- fa(surveys, fm="ml", rotate = "varimax", nfactors=3)

fit

corPlot(fit$r, numbers=TRUE, main="") # plot EFA

#####

# determine the reliability of the three factors #####

# Cronbach's alpha (internal item consistency)

# factor 1 -> items 4, 6a, 7, 7a, 7b, 8b

f1 <- c(4, 9, 11, 12, 13, 16)

factor_1 <- matrix(nrow = nrow(surveys), ncol = 0)

for(i in f1) {

  factor_1 <- cbind(factor_1, surveys[i])

}

```

```

factor_1
alpha(factor_1, check.keys=TRUE) # std.alpha = 0.75
# factor 2 -> items 2, 3, 6, 6b
f2 <- c(2, 3, 8, 10)
factor_2 <- matrix(nrow = nrow(surveys), ncol = 0)
for(i in f2) {
  factor_2 <- cbind(factor_2, surveys[i])
}
factor_2
alpha(factor_2, check.keys=TRUE) # std.alpha = 0.81
# factor 3 -> items 5, 5a, 5b, 9
f3 <- c(5, 6, 7, 17)
factor_3 <- matrix(nrow = nrow(surveys), ncol = 0)
for(i in f3) {
  factor_3 <- cbind(factor_3, surveys[i])
}
factor_3
alpha(factor_3, check.keys=TRUE) # std.alpha = 0.76
#####

```